Understanding the bug that makes our drugs: A multi-OMICs approach to engineer *Escherichia coli* for improved production of recombinant periplasmic proteins in the biopharmaceutical industry.

Alice Seleiro

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Medicine Dentistry and Health
School of Clinical Dentistry

26/09/2022

Abstract

Around 35% of therapeutic proteins produced in *E. coli* (the main bacterial expression system) are antibody fragments that are secreted to the periplasm [1]. However, optimizing expression requires trial and error and issues such as secretion system overloading and periplasmic inclusion body formation can occur.

In this thesis, proteomics and transcriptomics were used to analyse strains expressing a model therapeutic protein (an anti-TNF scFv) in the cytoplasm and periplasm under industrially relevant fermentation conditions. This data was then used to engineer strains and media that were tested for improved expression and secretion of periplasmic recombinant proteins.

ScFv secretion affected expression of multiple genes and proteins involved in stress (heat shock, envelope stress, iron stress and oxidative stress), cellular components (proteases and chaperones, the Sec secretion system and flagella) and metabolic pathways (oxidative phosphorylation, fatty acid, amino acid and ribosome synthesis). High levels of scFv expression (0.125 mM IPTG) that overloaded the secretion system also affected expression of the CpxR envelope stress response, pilus genes, osmotic stress, and membrane proteins.

Data from this study was used to develop and test strains altering expression of 18 genes and media components. Of these, 3 alterations to expression of genes involved in SecY and membrane protein degradation, improved scFv expression and secretion under shake flask conditions when inducing high levels of scFv expression (0.125 mM IPTG) that overloaded the secretion system. Initial data indicated that 2 of these strains improved yields of secreted scFv 134% and 156% (12 hours post-induction) under industrially relevant fermentations conditions and improved expression of different recombinant periplasmic proteins (HGH, sfGFP and Hel4).

Overall, this project has improved our understanding of cellular responses to recombinant protein secretion and has led to the development of novel methods to improve biologics production.

Acknowledgements

I would like to start my thesis by thanking my supervisors. I am extremely grateful for their input on this project and feel privileged to have worked with such great minds and people. It is also thanks to their feedback, reassurance, and encouragement that I have developed into the scientist I am today. Firstly, thank you to Prof. Graham Stafford for being supportive of me and the project, and always being there when I needed anything. My thanks also go to Dr Caroline Evans, for providing her proteomics and iTRAQ expertise, Prof. Jeff Green for providing his expertise in bacterial physiology and Chris Lennon, for his industrial input.

I also like to thank the BBSRC White Rose DTP for funding this project, FDB for in-kind contributions and the University of Sheffield for granting this project a Knowledge Exchange (KE) support grant to conduct RNA-seq.

Multiple people have helped me through this process and therefore I would also like to thank support staff at the University of Sheffield including Dr Emma Bird, technicians Brenka McCabe and Jason Heath for their support in running the labs and Chris Hill for TEM training. Many thanks also go to the molecular biology and microbial fermentation teams at FDB, including Christina Queenan, Dr Jeff McGeorge, Amy Woodhall, Deidre Boland and Danielle Fuller for their help preparing and running Ambr fermentations and Sarah Ryan for providing feedback on this project. Finally, I would like to thank Dr Rosemarie Gough for mentoring me through the thesis writing process.

I would also like to acknowledge my colleagues and friends at the Dental School for their camaraderie and encouragement. Particularly I would like to thank Ash, Beth, Kittie, Hollie, Jon, Caroline, Liz, Tom, Rachel and the rest of the Micro Group for their support.

My special thanks go to my amazing partner George for being extremely understanding and supportive of me and talking sense into me. I would also like to thank my family Ed, Mel and Tom for being amazing role models and being extremely supportive of everything I do.

Table of contents

List of abbreviations

2D-DIGE	Two-Dimensional Difference Gel Electrophoresis
2D-GE	Two-Dimensional Gel Electrophoresis
BDMA	Benzyldimethylamine
Bp	Base Pair
BSA	Bovine Serum Albumin
CID	Collision Induced Dissociation
Cyto	Cytoplasm
dAbs	Single Domain Antibody Fragments
DDSA	Dodecenyl Succinic Anhydride
EDTA	Ethylenediaminetetraacetic Acid
ESI	Electrospray ionisation
Fabs	Fragment Antigen Binding Site
FBA	Flux Balance Analysis
FDB	FUJIFILM Diosynth Biotechnologies
GO	gene ontology
HCDC	High Cell Density Cultures
Hel4	Heavy Chain Single Domain Antibody Fragment
HGH	Human Growth Hormone
IPTG	Isopropyl β-d1thiogalactopyranoside
iTRAQ	Isobaric Tags for Relative and Absolute Quantitation
LB	Luria-Bertani
LC	Liquid Chromatography
m/z	Mass to Charge Ratio
mAbs	Monoclonal Antibodies
MALDI	Matrix Assisted Laser Desorption Ionisation
MOBpsi	Multi-OMIC based production strain improvement
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
Nt	Nucleotides
OD600	Optical Density at 600 nm
P11	Paveway 11 Plasmid

Peri	Periplasm
PMF	Proton Motive Force
PPiase	Peptidyl-Proline Isomerase
RC	Reverse Complement
RNAP	RNA Polymerase
RPP	Recombinant Protein Production
ScFv	Single Chain Antibody Fragment
SDS	Sodium dodecyl sulphate
Sec	General Secretory Pathway
sfGFP	Super-Folder Green Fluorescent Protein
SID	Surface induced dissociation
SILAC	Stable Isotope Labelling with Amino Acids in Cell Culture
SOB	Super Optimal Broth
sp	Signal Peptide
SRP	Signal Recognition Particle
TAT	Twin-Arginine Translocation
TCEP	Tris Carboxy Ethyl Phosphene
TEAB	Tetraethylammonium Bromide
TEM	Transmission Electron Microscopy
TEMED	Tetramethylethylenediamine
TM	Annealing temperature
TMT	Tandem Mass Tagging
TNF	Tumour necrosis factor
TOF	Time of Flight
-ve	Negative

List of figures

List of Tables

List of Appendices

Chapter 1. Literature review

1.1 Recombinant protein production and secretion in *E. coli*

1.1.1 Background: Recombinant protein production and secretion in *E. coli*

1.1.1.1 Therapeutic protein production industry

Recombinant protein production (RPP) is the process of manipulating cells to overproduce a specific protein by introducing a gene of interest into the cell often using an expression plasmid. This has been an invaluable tool in multiple fields including fundamental research, synthetic biology and in industry to produce enzymes and therapeutic proteins.

Indeed, compared to extracting therapeutic proteins from biological samples, recombinant protein overexpression has improved yields and consistency of proteins, reduced costs and issues of disease transition and allowed for protein engineering.

The first therapeutic protein manufactured using this process was Somastin in 1977 [2] shortly followed by Insulin [3] and Human Growth Hormone (HGH) [4]. Since then, many other therapeutic proteins have been produced recombinantly including recombinant vaccines (e.g. Gardasil-HPV), enzymes, hormones and antibody therapies (e.g. Humira) with more than 316 biopharmaceuticals having been approved for use in the US and EU in 2018 [5]. Development of this technology and expertise has meant that recombinant protein production has become a large industry which was valued at 125.8 billion USD in 2020 and is growing with an estimated compound annual growth rate (CAGR) of 11.2% between 2021-2026 [6].

Despite this success, therapeutic proteins are some of the most expensive treatments on the market compared to small molecule drugs [7]. This is in part because therapeutic proteins are more complex and difficult to produce, often requiring lots of trial and error before a commercially viable process is developed [8], [9].

Methods used to overexpress recombinant proteins vary depending on their application. In research settings, tags are often used to improve protein solubility or aid purification. However, these tags are rarely used when producing therapeutic proteins as they can affect biological activity, cause adverse immunological effects and their removal is timely, costly and often leaves residues [10]. Furthermore, while recombinant proteins are often expressed in shake flasks in research settings, larger-scale High Cell Density Culture (HCDC) conditions are often used in industry. Bacteria are often grown in fed batch bioreactors where limiting nutrients (eg. carbon source, oxygen etc) are gradually added to fermenters. This allow bacteria to maintain growth and reach high densities (over 100 g cell weight/L) compared to batch fermentation cultures where no nutrients are added or continuous fermentations where media is both added and removed to maintain constant conditions [11], [12]. These high bacterial concentrations allow for higher yields, however it is also important to consider that this can add additional stress to cells and long fermentations can be more prone to mutations [12]–[14].

Multiple expression systems are used in industry to produce therapeutic proteins, these include mammalian systems (Chinese Hamster Ovary (CHO)), yeast (*Saccharomyces cerevisiae, Pichia pastoris*) and bacteria (*Escherichia coli*). In this section, we will specifically focus on the production of recombinant proteins in *E. coli*.

1.1.1.2 Therapeutic protein production in *E. coli*

E. coli is one of the most commonly used bacterial expression systems with around 30% of all therapeutic proteins on the market being produced in this bacterium [15]. Production of recombinant proteins in *E. coli* has multiple advantages [8], [16]. It is cheap as *E. coli* grows fast (with a doubling time of 20 mins under optimal conditions), up to high cell densities, in cheap media [11], [13]. Furthermore, *E. coli* is a model organism that is well characterised, has a relatively small genome [17] and has multiple tools available to manipulate it such as plasmids and recombineering methods [18].

However, there are drawbacks to overexpressing recombinant proteins in *E. coli* which many studies have tried to overcome through strain engineering

[19]–[21]. Firstly, *E. coli* produces LPS which is toxic when injected into humans and therefore proteins produced in this organism require a high level of purification. Secondly, *E. coli* produces acetate which can accumulate under High Cell Density Culture (HCDC) negatively affecting bacterial growth and protein yields [12]–[14], [22]. Thirdly, unlike mammalian and yeast expression systems, *E. coli* does not naturally glycosylate proteins. Proteins can be glycosylated via N- or O-linked glycosylation of specific residues. This glycosylation plays an important role in protein function and immunogenicity and 70% of all therapeutic proteins [23] including novel therapies such as viral vectors and monoclonal antibodies (mAbs) are glycosylated. Strains have been engineered to produce glycosylated proteins however these strains often do not produce "human like" glycosylation patterns and are not currently economically viable [21], [24]. Therefore, mammalian and yeast systems) currently dominate the recombinant protein production industry [1], [25]. However, one area where *E. coli* remains relevant is in the production of antibody fragments [25] which represent >35% of all therapeutic proteins produced in *E. coli* in 2018 [1].

1.1.1.3 Antibody fragments

Antibody fragments are segments of monoclonal antibodies (mAbs) containing the variable antigen binding site (Figure 1). There are a wide variety of antibody fragments of different sizes and structures. The most common types of antibody fragments are single-chain variable (scFv), single-domain antibody (dAbs) and fragment antigen binding (Fab) fragments (Figure 1). These antibody fragments are either derived through digestion (eg. Fabs) or engineering (eg. scFvs where a linked is added between the heavy and light chain of the variable regions). Antibody fragments can be used to treat cancers and immune diseases, for targeted drug delivery and diagnostics (Table 1) [26]–[28]. Antibody fragments have several advantages including improved tissue penetration (due to their size) and lower immunogenicity, however, they have a shorter half-life in the body compared to full length antibodies [26], [28]. Antibody fragments are often cheaper to produce as they do not need to be glycosylated and can therefore be produced in bacterial systems such as *E. coli.* However, antibody fragments contain disulphide bonds with scFvs containing 1-2 disulphide bonds

and Fabs containing 5 disulphide bonds. Antibody fragments therefore need to be secreted to the periplasm.

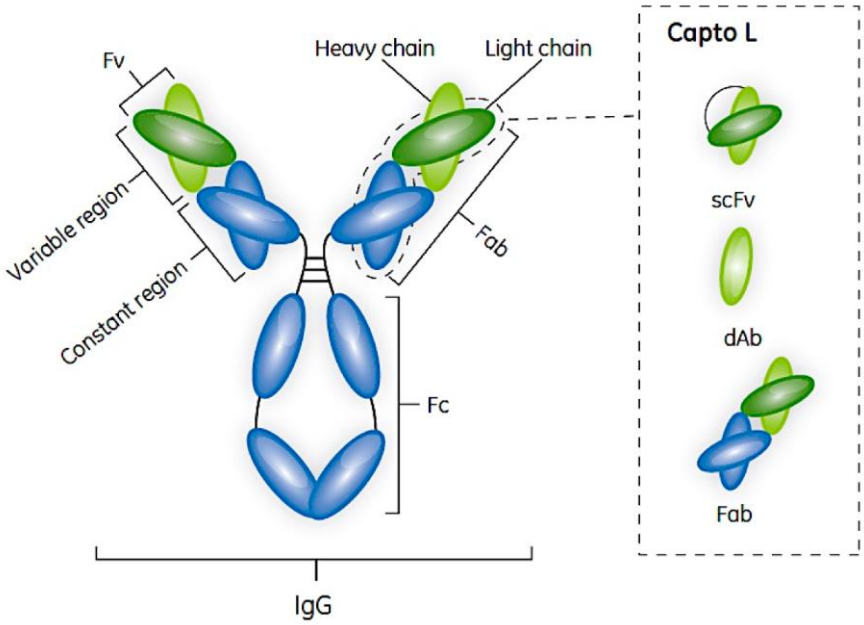


Figure 1. Full-length antibody (IgG) and the most common antibody fragments. Antibody fragments include a single-chain variable fragment (scFv), single-domain antibody fragment (dAbs) and fragment antigen binding fragment (Fab). The constant region is indicated in blue and the variable antigen binding region in green. The scFv also contains a linker between the heavy and light chains as indicated by a line. Adapted with permission from [29].

Drug name	Type of antibody fragment	Application
Abciximab	Chimeric Fab	Blood clotting disorder treatment
Axicabtagene Ciloleucel	scFv	Lymphoma treatment
Blinatumomab	scFv	Leukaemia treatment
Brolucizumab	scFv	Neovascular age-related macular degeneration treatment
Caplacizumab	dAb	Thrombocytopenic purpura treatment
Certolizumab pegol	Fab	Crohn's disease treatment
Idarucizumab	Fab	Anticoagulant
Moxetumomab pasudotox	scFv conjugated to pseudomonas toxin	Leukaemia treatment
Nofetumomab Merpentan	Fab	Diagnostic small-cell lung cancer
Ranibizumab	Fab	Neovascular age-related macular degeneration treatment
Tisagenlecleucel	scFv	Leukaemia treatment

Table 1. List of recombinant antibody fragments approved for use by the FDA and their applications. These include single-chain variable fragments (scFv), single-domain antibody fragments (dAbs) and fragment antigen binding fragments (Fabs). Information is derived from [26], [28]**.**

1.1.1.4 Secretion of therapeutic proteins to the periplasm in *E. coli*

Multiple therapeutic proteins including antibody fragments and hormones (eg. HGH) are secreted to the periplasm (the area between the inner and outer membrane) in *E. coli*. As mentioned previously, recombinant proteins are often targeted to the periplasm if they contain disulphide bonds. Disulphide bonds are formed when two cysteine residues are oxidised and play an important role in protein folding and function. Unlike the cytoplasm, the periplasm is an oxidising environment and therefore disulphide bonds can form in this cellular compartment. Furthermore, the periplasm contains proteins that aid disulphide bond formation DsbABCDG [30], [31] (For further detail about Dsb proteins see section 1.1.2.1.3). Strains have been engineered to produce

disulphide bond-containing proteins in the cytoplasm by mutating endogenous reducing pathways, overexpressing oxidases and dsb proteins such as CyDisCo [32]–[35], Shuffle [36] and Origami strains [37]. However, most disulphide bond containing proteins are still expressed in the periplasm.

Overexpressing recombinant proteins in the periplasm has other advantages including increasing protein stability, reducing protein degradation and reducing cytoplasmic toxicity of certain proteins [38]–[40]. Secreted recombinant proteins are also more representative of those found in nature as they have a true N-terminus without a methionine [26]. Furthermore, protein purification can be more straightforward when taking this approach as overexpressed proteins can be extracted from cellular fractions that are a lot less complex such as periplasmic fractions extracted using osmotic shock [41] or the supernatant due to cell leakiness when overexpressing periplasmic proteins [42], which is the most common method used to extract proteins in industry.

In the following sections key pathways involved in expression of recombinant proteins in the periplasm, issues that can arise from periplasmic expression and current methods that have been developed to overcome these issues will be discussed.

1.1.2 Key players in recombinant protein secretion

Secretion of recombinant proteins involves several key steps: directing proteins to the secretion system, secreting proteins, and folding proteins in the periplasm. The key processes involved in these steps will be discussed in the following sections.

1.1.2.1 Secretion systems

Recombinant proteins are directed to the periplasm by adding an N-terminal signal peptide which is cleaved off by signal peptidases once the protein is translocated to the periplasm [43]. Two main secretion systems are used to secrete recombinant proteins to the periplasm: the TAT (Twin-Arginine Translocation) system which secretes folded proteins [33], [44] and the Sec (general secretory pathway) system which secretes unfolded proteins [39], [45],

[46] (Figure 2). These secretion systems will be discussed in the following sections.

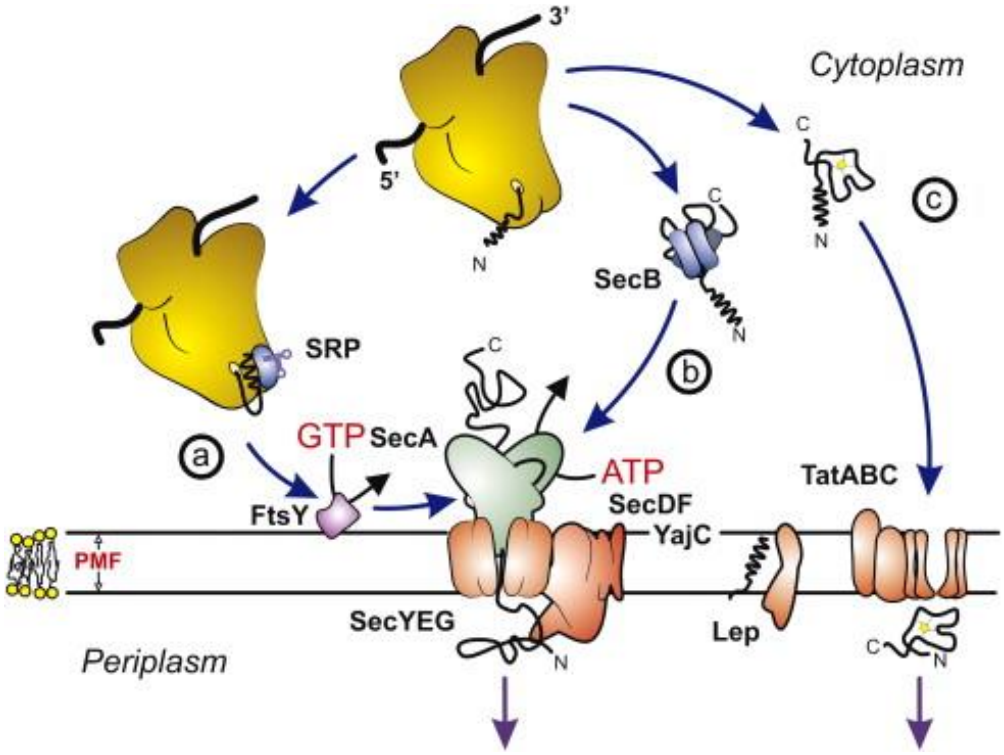


Figure 2. Different types of secretion pathways used to secrete recombinant proteins into the periplasm. (a) co-translational and (b) post translational Sec based secretion and (c) TAT based secretion. Figure is adapted with permission from [45]***.***

1.1.2.1.1 The Tat secretion system

Recombinant proteins have successfully been secreted to the periplasm using the TAT (Twin-Arginine Translocation) secretion pathway [33], [44], [47]–[50].

Recombinant proteins are directed to the TAT secretion system by adding an n-terminal TAT signal peptide. TAT signal peptides can be distinguished by their long n-terminus, TAT consensus motif, less hydrophobic core and more positively charged c-terminus (Figure 3) [45], [51].

This signal peptide binds to membrane proteins TatBC. Once bound, TatBC forms a complex with TatABC which translocates folded proteins across the inner membrane (Figure 2c). TAT secretion is ATP independent and relies on proton motive force (PMF) [45], [52]. However, the mechanism by which TatA translocates proteins is contested with the two main theories being that it forms a pore in the membrane or depolarises the membrane [45], [52].

The main advantage of using the TAT secretion system to secrete recombinant proteins is that it only secretes correctly folded proteins improving quality control and purification [48], [50]. There are however drawbacks to using this system. TAT secretion is not suitable for the production of proteins containing disulphide bonds unless specialised strains such as CyDisCo are used [33]. Furthermore, TAT secretion is often less efficient than Sec secretion [39], [53].

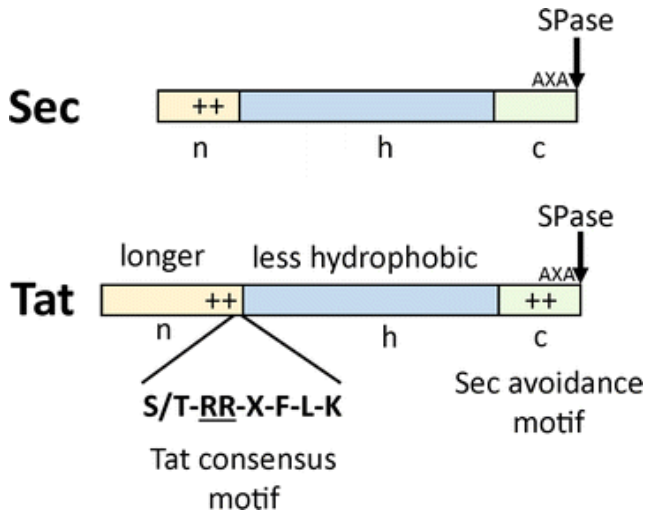


Figure 3. Figure representing similarities and differences between sec and tat signal peptides. N represents the positively charged n-terminus, h represents the hydrophobic h-region and c represents the c terminus containing a signal peptidase cleavage sequence (A-X-A motif). Figure adapted with permission from [51]***.***

1.1.2.1.2 The Sec secretion system

The Sec (general secretory pathway) system is one of the most common secretion systems used to secrete recombinant proteins to the periplasm [39] and is used in this study. It can be used to insert proteins into the inner membrane with the aid of YidC however this will not be covered in this section. The Sec secretion system is ATP dependent and is aided by Proton Motive Force (PMF) with the help of the membrane proteins SecDF [43]. Proton Motive Force (PMF) is however not essential for Sec secretion [54]. For an in-depth review of the Sec secretion system see [43], [55], [56].

The Sec secretion system can secrete proteins via two main pathways: SRP-dependent or SecA-dependent secretion. The pathway used depends on the properties of the N-terminal signal peptide with SRP-dependent signal peptides being more hydrophobic [39].

1.1.2.1.2.1 SRP-dependent secretion

In the SRP-dependent pathway, proteins are mostly secreted co-translationally where translation and secretion occur simultaneously (Figure 2a). Signal recognition particle (SRP) binds to signal peptides once they emerge from the ribosome and stalls further translation. SRP then binds to its receptor FtsY in the inner membrane which through GTP hydrolysis transfers the ribosome to the SecYEG translocon [55]. The binding of FtsY and the ribosome to SecYEG changes the conformation of SecYEG activating the channel. Translation of the protein then allows for translocation of the protein through the secretion system [43]. The SRP-dependent pathway is mainly used to insert inner membrane proteins however there are exceptions to this rule.

1.1.2.1.2.2 SecA-dependent secretion

The SecA-dependent secretory pathway primarily targets proteins for secretion post-translationally and is generally used to secrete proteins to the periplasm (Figure 2b).

SecA guides preproteins to the sec secretion system and is often found binding to the ribosome or the SecYEG complex [57]. SecB can aid this secretion by guiding preproteins to SecA, however is not essential for secretion [58]–[60]. SecA then binds and changes the conformation of SecY and activates the SecYEG channel. SecA uses it's ATPase activity to translocate proteins through the pore. Proteins are thought to be translocated either through a "push and slide" mechanism where SecA pushes the protein through the pore [61] or through Brownian movement where proteins are transported by protein folding in the periplasm [62].

Cytoplasmic chaperones and proteases, such as SecB, Tig, DnaK/DnaK and GroEL/GroeES also play an important role in post-translational secretion by preventing folding of preproteins in the cytoplasm and degrading incorrectly folded proteins [63], [64].

1.1.2.1.3 Protein folding in the periplasm

Sec secreted proteins then need to be correctly folded in the periplasm. Multiple proteins play a role in periplasmic protein folding and often have

overlapping functions (Figure 4) [46], [65]–[69]. These include chaperones that help fold proteins (e.g., Skp, Spy etc…), Dsb proteins that aid in disulphide bond formation, peptidyl-prolyl isomerases (PPIases), which transforms the conformation of cis and trans peptide bonds and proteases that degrade incorrectly folded proteins.

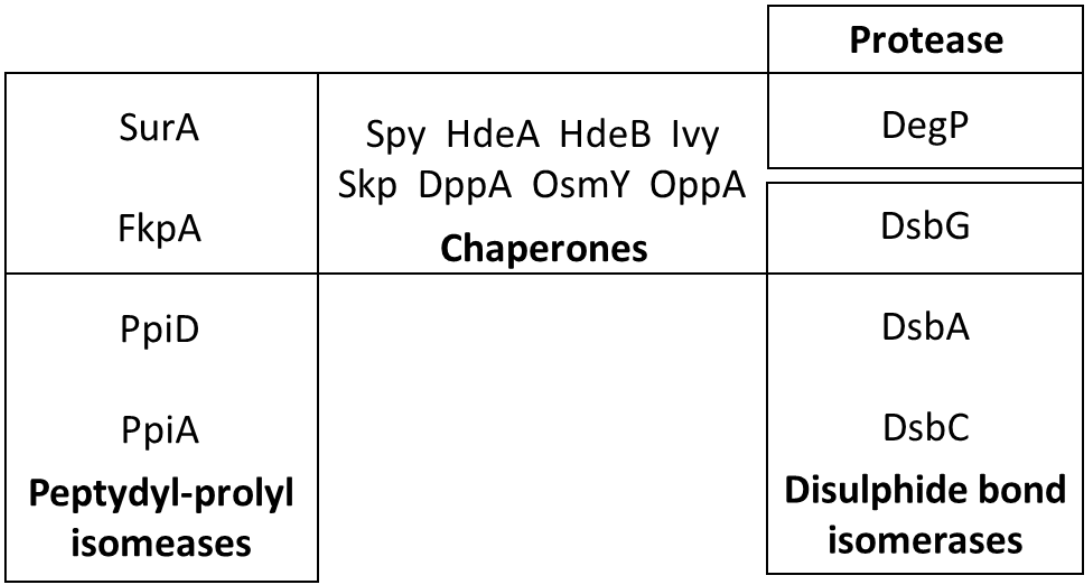


Figure 4. Proteins involved in periplasmic protein folding in E. coli. These include prolyl isomerases (PPIases), chaperone like protein, proteases and disulphide bond isomerases. Figure is an updated version of a figure from [65]***.***

Multiple periplasmic chaperones aid in protein folding (Figure 4). Unlike cytoplasmic chaperones, periplasmic chaperones fold proteins in an ATP independent fashion as the periplasm is devoid of ATP [46], [65]–[69]. Periplasmic chaperones are therefore often described as "chaperone-like proteins" and bind to preproteins to direct their folding. Because of this, it was long thought chaperones did not exist in the periplasm until their relatively recent discovery [68], [70]. The activity of these chaperone like proteins has been confirmed both in vitro (eg. overexpression and knockout strains) and in vivo (e.g. aggregation assays) [70], [71]. Many of these periplasmic chaperones are known to respond to specific stimuli such as HdeA and HdeB responding to acid stress, OsmY responding to osmotic stress and with DppA and OppA also playing an important role in transport of peptides via ABC transporters.

Dsb proteins also play a role in the formation of disulphide bonds (Figure 5) [72], [73]. DsbA catalyses disulphide bond formation in the periplasm [74]. In

doing so, cysteines in DsbA become reduced and can be restored by the inner membrane proteins DsbB which itself is oxidised by ubiquinone or menaquinone [75]. However, disulphide bonds can form between the wrong cysteines. DsbC and DsbG, therefore, recognise misfolded proteins and act as disulphide bond isomerases to rearrange the disulphide bonds. In doing so DscC and DsbG become oxidised and therefore need to be reduced by the inner membrane protein DsbD to remain active.

Expression of periplasmic protein folding genes is highly regulated by the envelope stress response which will be covered in the following section.

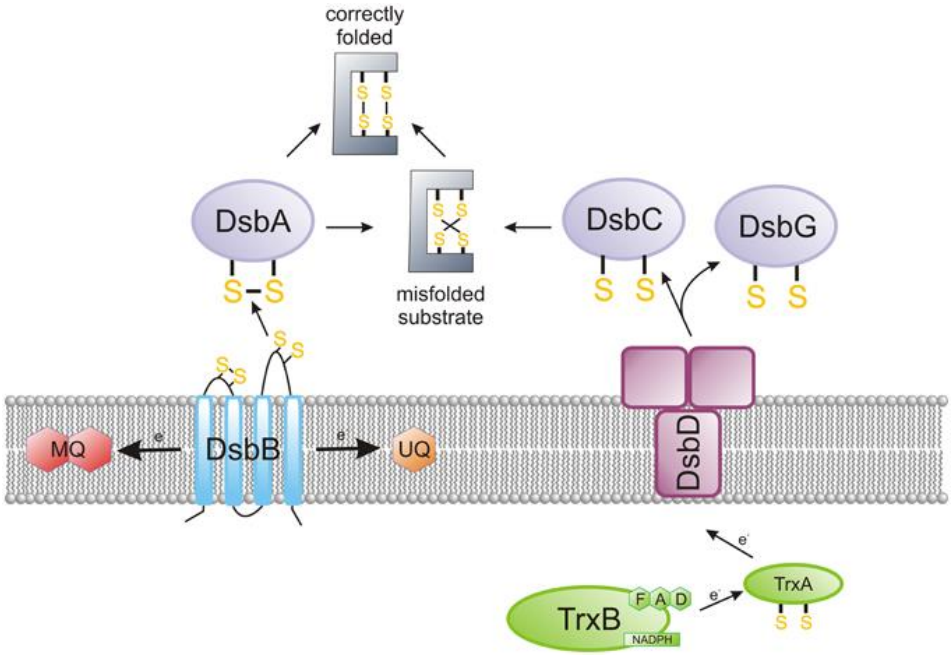


Figure 5. Proteins involved in disulphide bond formation in E. coli. DsbA is present in the periplasm and catalyses disulphide bond formation. The inner membrane protein DsbB is oxidised by ubiquinone or menaquinone and restores DsbA to its oxidated state. DsbC and DsbC are isomerase that ensure that the correct disulphide bonds are formed. Thioredoxin reductase reduces DsbB which in turn restores and reduces DsbC and DsbG. Figure adapted with permission from [76]***.***

1.1.2.1.4 Envelope stress response

The envelope stress response is similar to the heat shock response but also regulates expression of genes involved in periplasmic and membrane protein folding [77]–[79]. The extracellular stress response is activated by multiple stimuli that cause membrane damage or lead to periplasmic and membrane protein misfolding [79], [80]. There are multiple overlapping but

distinct regulators of the envelope stress response: the SigmaE, Cpx, Rcs, Bae and Psp pathways [65], [66], [69], [77]–[82].

The SigmaE/RpoE pathway is activated by binding of unfolded membrane proteins to DegS which triggers a cleavage cascade resulting in the release of RpoE from the anti-sigma factor RseA allowing the sigma factor to bind to RNA polymerase (RNAP) activating expression of several genes [77], [83].

The Cpx, Rcs and Bae pathways are regulated by two component systems. These are composed of a membrane bound histidine kinase sensor (CpxA, RcsD or BaeS) which upon activation phosphorylates activates a regulator (CpxR, RcsB (on its own or dimerised with RcsB or RcsA) or BaeR) which activates or represses expression of several genes [84].

The Psp response is regulated by PspF which activates expression of psp genes. PspF is sequestered by PspA when it is not bound to PspB and PspC [85], [86]. PspA is the main effector of this system and plays a role in maintaining proton motive force [87], [88].

Both the Sigma E and Cpx pathways have been well characterised, however less is known about Rcs, Bae and Psp pathways [65], [78], [81].

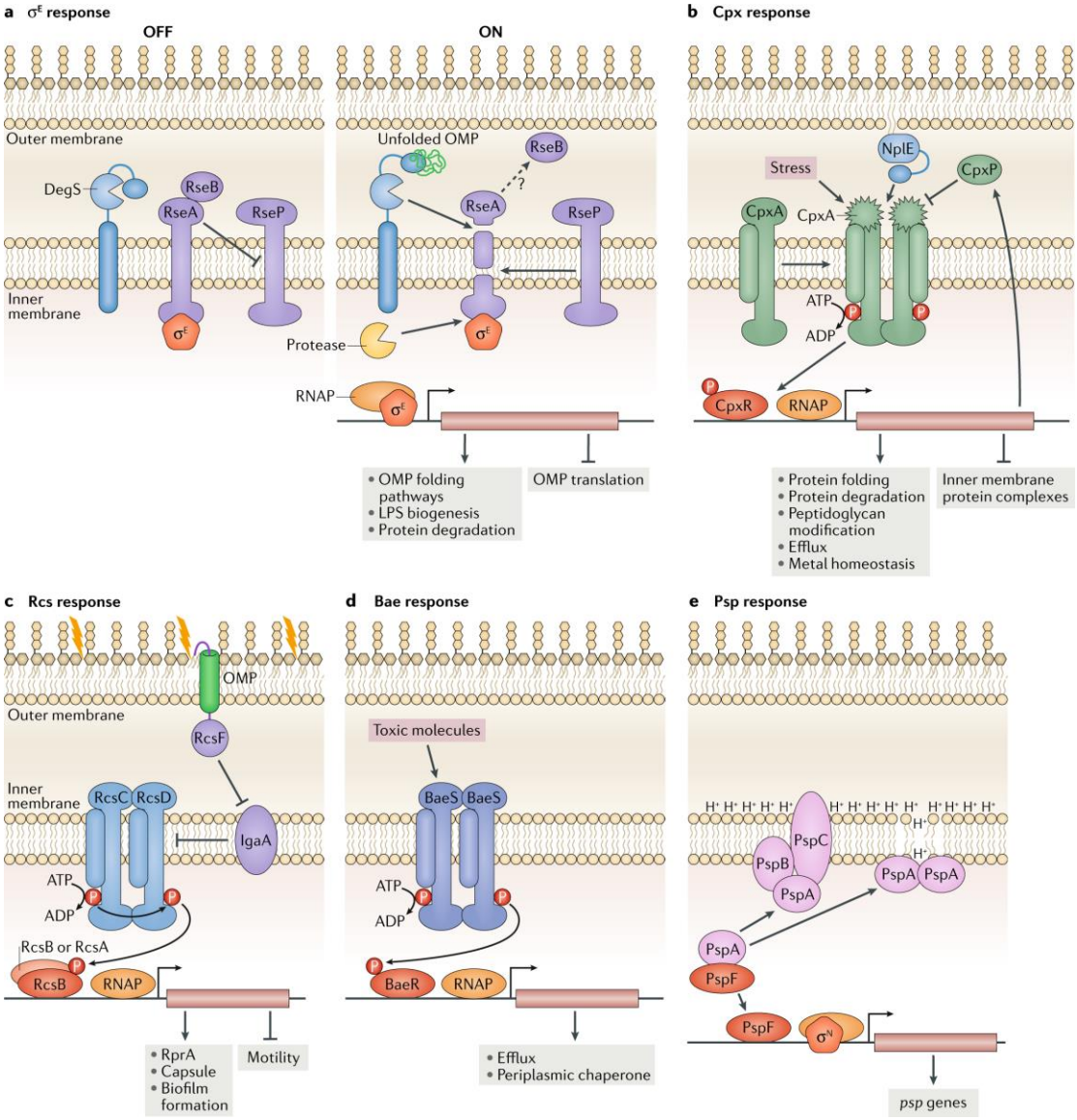


Figure 6. Figure representing envelope stress response pathways: Sigma E, Cpx, Rcs, Bae and Psp pathways. Key effectors, regulators and regulons are indicated. (a)The Sigma E response is activated by binding of unfolded outer membrane proteins (OMP) to DegS which triggers a cleavage cascade resulting in the release if Sigma E from the anti-sigma factor RseA. Sigma E then binds to RNA polymerase (RNAP) regulating expression of genes involved in outer membrane protein folding. (b) The Cpx response is activated by multiple stresses including overexpression of NplE leading to activation and phosphorylation of CpxR by CpxA and regulation of multiple genes. This stress response is negatively regulated by CpxP. (c) The Rsc response is activated by the RcsF sensor binding to the negative inhibitor IgaA, which activates the RcsCD phosphorelay system that in turn phosphorylates RscB. RscB then activates transcription alone or in a dimer with RscB or RscA. (d) The Bae response is activated by the BaeS histidine kinase which phosphorylates and activates BaeR regulating gene expression. (e) The Psp response is activated by binding of PspA to the PspBC membrane complex, releasing PspF which activates expression of the Psp operon. PspA also binds to the membrane preventing proton leakage. Figure was adapted with permission from [82].

1.1.3 Issues when secreting recombinant proteins to the periplasm in *E. coli*

Multiple issues can arise when overexpressing recombinant proteins that lead to low yields such as overwhelming the protein production and folding machinery as well as recombinant proteins being toxic to the cell. There are however specific issues that can arise when trying to secrete recombinant proteins to the periplasm. Firstly, the periplasm is limited in space, making up around 8–16% of the total cell volume [67]. The quantity of proteins that can be expressed in the periplasm is therefore greatly reduced compared to the cytoplasm [89], [90]. Furthermore, overexpression of large amounts of recombinant proteins in the periplasm can lead to cell leakiness and lysis [42], [91]. Leakiness is however sometimes desirable as this helps with downstream processing as proteins can be extracted directly from the supernatant [42].

Another issue can be incorrect processing of recombinant protein in the periplasm. Sometimes, signal peptides are improperly processed and cleaved [92]. Furthermore proteins do not always fold properly in the periplasm leading to the formation of periplasmic inclusion bodies [66], [93]–[95]. This can be due to multiple factors including the property of the protein overexpressed and periplasmic protein folding systems getting overwhelmed. Periplasmic inclusion bodies can be toxic to the cell [96] and are often no longer functional and need refolding.

High levels of recombinant protein secretion have been shown to overwhelm the secretion system leading to accumulation of the recombinant protein as well as other native membrane and periplasmic proteins in the cytoplasm [63], [89], [90], [97], [98]. This secretion system overloading leads to a decrease in the yields of secreted recombinant proteins. It can also have a negative impact on the cell, as native membrane and periplasmic proteins aren't properly exported [89], [90], [97], [98]. Furthermore, unsecreted proteins often do not fold properly and form cytoplasmic inclusion bodies which can be toxic to the cell.

1.1.4 Methods of improving recombinant protein secretion to the periplasm in *E. coli*

Multiple methods have been developed to overcome issues with recombinant protein secretion to the periplasm so that yields are improved. These include optimizing signal peptides, controlling protein expression levels and genetically engineering strains. However, no universal method has been developed to improve expression of all periplasmic recombinant proteins [99]. Because of this multiple strains need to be screened to identify the strain that improves the production of a specific protein [100], [101]. For an in depth review of methods developed to improve secretion see [102], [103].

1.1.4.1 Signal peptides

Screening different signal peptides is one of the most common ways of optimizing recombinant protein secretion [98], [104]. Recombinant protein production and secretion levels have been shown to drastically change depending on which signal peptide us used [15], [105]–[108] and there is no current method of predicting which signal peptide is best to secrete a specific protein.

Signal peptides are often identified from highly secreted proteins (eg. OmpA, PhoA, DsbA and Hbp signal peptides). Some studies have been able to further improve recombinant protein production and secretion by engineering signal peptides using mutagenesis screens [15], [63], [105], [109]–[111]. This has also enabled us to understand key determinants of an efficient signal peptide including the hydrophobicity and charge of the h-region.

1.1.4.2 Controlling expression levels

Another important method of optimizing recombinant protein secretion is tuning expression levels so that the secretion system is not overloaded. This has been done by testing recombinant protein secretion when adding different amounts of inducer [90], [98], [112], [113], screening different promoters [114] and plasmids with different copy numbers [113], [115]. This has also been done by controlling translation of proteins using riboswitches [116] or mutating

translation initiation sites [117]. Controlling expression levels can also be simultaneously combined with screening signal peptides to even further increase periplasmic secreted protein yields [98].

1.1.4.3 Genetically engineering strains to improve recombinant protein secretion

Genetic engineering has been an important tool in improving recombinant protein production [118]–[121]. The following section will specifically focus on how genetic engineering has been applied to improve recombinant protein secretion to the periplasm.

There are two main approaches to engineering strains: targeted methods that involve overexpressed or knocked out genes based on our current knowledge or global methods that involve screening mutants or evolving strains to improve protein production. Targeted methods however rely heavily on trial and error and prior knowledge about gene function. Global methods also have the disadvantage that they require methods to screen high yields of recombinant protein production and secretion. Furthermore, it can often be challenging to determine which mutation is responsible for a change in phenotype [119], [122]–[124].

Some studies have successfully used global approaches including evolution and mutagenesis to improve recombinant protein secretion [53], [101] however the vast majority of studies take targeted approaches which will be covered in the following section.

1.1.4.3.1 Genetic engineering - overexpressing cytoplasmic chaperones

As mentioned previously in section 1.1.2.1.2.2, cytoplasmic chaperones play an important role in stabilizing proteins in the cytoplasm, preventing their folding and guiding proteins to the appropriate secretion system. Studies have found that overexpressing cytoplasmic chaperones SecA and SecB [125], DnaK/DnaJ [101], [126]–[129] and GroEL/GroES [101] improves recombinant protein production and secretion. Knocking out the cytoplasmic chaperone Tig (also known as trigger factor) has also been shown to improve recombinant

protein secretion as this chaperone is thought to slow down protein secretion [130], [131].

1.1.4.3.2 Genetic engineering - overexpressing the secretion system

Another method for improving recombinant protein secretion involves increasing the cell's secretion capability by co-expressing the secretion system used. This approach has been particularly useful in improving the quantity of secreted protein via the TAT secretion pathway, particularly as this expression system is expressed in low quantities in the cell. Overexpressing the TatABCD translocase improved secretion of GFP and HGH 2-5 fold when using a signal peptide that directed proteins to the Tat translocon [132], [133]. Overexpressing the Sec secretion system has however shown inconsistent effects on recombinant protein production however some studies have found that overexpressing SecYE or mutating SecY (prlA4) improved yield of secreted proteins [101], [134].

1.1.4.3.3 Genetic engineering - overexpressing signal peptidases

Signal peptidases play an important role in cleaving off signal peptides once proteins arrive in the periplasm. However, sometimes signal peptides aren't properly processed. Studies have therefore found that co-expressing signal peptidases such as SppA improved production of an scFv antibody fragment (AP) and Granulocyte-macrophage colony-stimulating factor (GM-GCF) 1.7 and 2.5 fold when secreted to the periplasm [99].

1.1.4.3.4 Genetic engineering - overexpressing genes involved in periplasmic protein folding

Overexpressing genes involved in periplasmic protein folding has also been shown to improve recombinant protein secretion. Extensive research has shown that overexpressing Dsb proteins, particularly DsbA and DsbC improved production of certain disulphide bond containing proteins in the periplasm [99], [101], [127], [135]–[137]. Other studies have also found that co-expressing the

periplasmic chaperone Skp or the PPIase FkpA improved production of recombinant proteins in the periplasm [100], [101], [138]–[143]. Inhibiting protein degradation by knocking out periplasmic proteases such as DegP, has also improved yields of secreted recombinant proteins [144], [145].

Many of these proteins involved in periplasmic protein folding are regulated by the envelope stress response which responds to unfolded periplasmic and membrane proteins (see section 1.1.2.1.3). Therefore, studies have found that altering the envelope stress response such as overexpressing PspA, improved recombinant protein secretion to the periplasm when secreting protein via Sec and Tat pathways [146]–[149].

1.2 A systems biology approach to improving recombinant protein production in *E. coli*

1.2.1 Why OMICs?

Ever since the development of recombinant protein production as a technique, multiple studies observed that both plasmid presence and protein overexpression can have a negative impact on growth and protein yields [150]–[155]. It was therefore hypothesised that protein overexpression and plasmid maintenance placed a “metabolic burden” on the cell [150], [156]. Other studies also observed that recombinant protein overexpression activated multiple stress responses [157], as seen by the increase in expression of genes involved in the heat shock response when expressing insoluble proteins that form inclusion bodies [158], [159].

There has therefore been growing interest in understanding cellular responses to recombinant protein production with the hope of developing novel strains and strategies to improve recombinant protein overexpression by overcoming bottlenecks, alleviating stress and altering expression of key genes [160], [161].

The development of OMICs technologies (transcriptomics, proteomics and metabolomics) that can analyse levels of multiple transcripts, proteins and metabolites simultaneously has drastically increased our understanding of cell physiology, providing an less biased view of cellular responses and providing a

snapshot of what is happening in the cell. OMICs data also has the advantage that it can be integrated into metabolic models [162], [163] where flux balance analysis (FBA) and systems biology approaches can be used to identify bottlenecks [162]–[164]. Furthermore, a multi-OMICs approach can be combined to present a more comprehensive view of cellular responses [165].

This section will cover the use of OMICs in the field of recombinant protein production. Different OMICs technologies will be covered in the first section focusing specifically on techniques used in this study (iTRAQ and RNA-seq). Then studies that have previously used OMICs analysis to improve our understanding of recombinant protein production in *E. coli* will be discussed, with a focus on recombinant protein secretion to the periplasm.

1.2.2 OMICs technologies

1.2.2.1 Transcriptomics

1.2.2.1.1 Transcriptomics definition

Transcriptomics is the study of the transcriptome, which involves determining the level of expression of transcripts (RNA) within a sample. These transcripts can either be coding RNA (mRNA) that can be translated into proteins or non-coding RNA (ncRNA) that is known to have multiple functions including regulating gene expression. Two main methods are used to analyse the transcriptome: microarrays and RNA-seq.

1.2.2.1.2 Microarrays

Initial transcriptomics studies used microarrays to analyse the transcriptome [166]. Microarrays rely on the sequence-specific hybridization of cDNA to gene specific oligonucleotide probes that are spotted onto a glass slide. Microarrays allow for relative quantification of mRNA by labelling samples with different fluorescent stains and comparing hybridization on the same array or across multiple arrays. One of the main issues with microarrays is their lack of accuracy due to non-specific hybridization [167] and issues with dynamic range which can make it difficult to quantify low abundance transcripts. Furthermore, microarrays can only be conducted for organisms that have had their genome

sequenced and rely on accurate annotation of coding regions. Also, many initial microarrays only quantified a portion of the transcriptome [168]–[170].

1.2.2.1.3 RNA-seq

RNA-seq has now surpassed microarrays as the main technique used to study the transcriptome [171]. RNA-seq uses next generation sequencing to sequence transcripts (cDNA or RNA). Absolute or relative transcript expression is then determined by mapping these sequences to determine the number of reads that map to a specific gene. This allows for a near complete analysis of the entire transcriptome so long as sufficient sequencing depth is used and ribosomal RNA has been depleted [172].

RNA sequencing (RNA-seq) presents several advantages over microarrays. RNA-seq is a highly accurate method for determining transcript abundance as it doesn't have issues with dynamic range and non-specific hybridisation [173]. Furthermore, RNA-seq can be used to analyse organisms that do not have their genome sequenced, presents less bias and can be used to detect post transcriptional modifications in eukaryotes [173]. The cost of next generation sequencing technologies has decreased in the last 20 years making RNA-seq more accessible, however, there is still a large requirement for computational power associated with RNA-seq which can be costly [171].

Multiple next generation sequencing technologies have been used for RNA-seq. These can be grouped into two main categories, second generation technologies such as Ion Torrent and Illumina which sequence shorter sections of cDNA (up to around 300 bp [174]) and third generation sequencing such as PacBio SMRT sequencing and Nanopore sequencing which can sequence longer cDNA fragments (50 kb or longer [174]) and sequence RNA directly. Long-read sequencing has several advantages including improving mapping and detection of isoforms, however, methods such as Nanopore sequencing often have high error rates [175]. Direct RNA sequencing also presents several advantages including avoiding amplification biases [176].

1.2.2.1.4 Conclusion - Transcriptomics

While transcriptomics provides an insight into transcript expression, mRNA levels often do not correlate with protein expression levels due to differences in transcript and protein stability and differences in translation rates [177]. Proteomics has therefore played a critical role in understanding cell physiology and will be covered in the following section.

1.2.2.2 Proteomics

1.2.2.2.1 Proteomics definition

Proteomics involves analysing the proteome by comparing the expression of proteins between samples. Proteins play a critical role in cells and therefore analysing changes in protein abundance is important to understanding cellular physiology. Furthermore, the subcellular localisation of proteins can be analysed by fractionating samples. However, due to the complexity of the proteome and limitations of mass spectrometry, proteomics studies often only quantify about half of the *E. coli* proteome [178]. Despite limitations in coverage, proteomics is still widely used. Multiple methods have been developed to analyse the proteome and will be covered in the following sections with a particular focus on the method used as part of this study (iTRAQ). A summary of the advantages and disadvantages of these methods can be found in Table 2.

	Advantages	Disadvantages
Gel based methods		
2D-GE	Visualise post-translational modifications	Gel to gel variation Low throughput Requires good gel resolution Requires lots of replicates Low sensitivity Issues with dynamic range
2D-DIGE	Visualise post-translational modifications Improved dynamic range Multiplexing (max 3 tags) Avoids inter-gel variation	Requires good gel resolution Low sensitivity Issues with dynamic range
Gel-free methods		
Label free	Does not require expensive tags	Variation LC-MS/MS runs Requires many repeats Long run times
SILAC	Reduces technical variation in sample preparation Multiplexing (3 tags) Accurate quantification	Only suitable for specific cell types/organisms Expensive tags
TMT	Reduces technical variation in sample preparation Multiplexing (max 18 tags) Reduced spectral complexity	Underestimates differences in protein abundance
iTRAQ	Reduces technical variation in sample preparation Multiplexing (max 8 tags) Reduced spectral complexity	Underestimates differences in protein abundance

Table 2. Main advantages and disadvantages of proteomics technologies used for relative quantification. 2D-GE – Two-Dimensional Gel Electrophoresis, 2D-DIGE – Two-Dimensional Difference Gel Electrophoresis, iTRAQ – Isobaric Tags for Relative and Absolute Quantitation, TMT - Tandem Mass Tagging and SILAC - Stable Isotope Labelling with Amino Acids in Cell Culture.

1.2.2.2.2 Evolution of proteomics tools

1.2.2.2.2.1 Gel-based methods

Initial proteomics studies used gel-based methods to study the proteome. Proteins are separated using 2D gel electrophoresis (2d-GE) based on their molecular weight (MW) and isoelectric point (pI) (Figure 7). The intensity of

spots is then compared between samples and proteins that are differentially expressed are cut out and identified using mass spectrometry. Software can also be used to predict each spot based on its molecular weight and isoelectric point [179].

This method does however have multiple disadvantages (Table 2). Multiple replicates of each gel are needed as there can be technical variation between gels. This technical variation has been overcome by methods such as difference gel electrophoresis (DIGE) where samples are labelled with up to three fluorescent labels and run on a gel simultaneously, however, these tags are expensive [180]. Gel based methods also have the disadvantage that they are low throughput, have issues with dynamic range, are not good at quantifying low abundance proteins and require good resolution to separate all proteins [179].

Gel based methods do however present advantages when looking at post-translational modifications such as glycosylation and phosphorylation. Proteins with different post-translational modifications are often separated on a 2D gel and specific post-translational modifications can be stained for [179]. This is particularly important as post translational modifications often play an important role in protein activity.

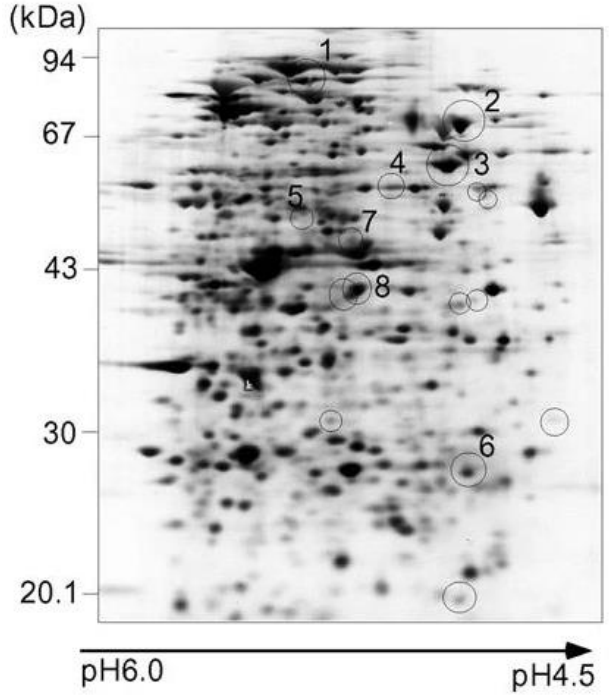


Figure 7. Example of a 2D gel. Proteins are separated based on molecular weight (horizontal) and isoelectric point (vertical). Adapted with permission from [181]***.***

1.2.2.2.2.2 Gel-free methods

Unlike gel-based methods, gel-free proteomics methods quantify and identify proteins simultaneously using tandem mass spectrometry (MS/MS). This section will start by briefly describing tandem mass spectrometry and go on to describe different gel-free quantification methods.

1.2.2.2.2.2.1 Principles of mass spectrometry

Gel free methods rely heavily on tandem mass spectrometry. Proteins are digested into peptides, most commonly with trypsin, in what is known as a bottom-up approach. These peptides are then loaded onto a tandem mass spectrometer that is often coupled with HPLC to further separate peptides. During tandem mass spectrometry peptides are ionised and separated based on mass to charge ratio (m/z) in MS1 (Figure 8). These ions are then fragmented and further separated based on mass to charge ratio (m/z) in MS2 before being detected (Figure 8). This spectrum can then be used to sequence the peptides based on peptide fragmentation patterns. However, mass spectrometry does not provide absolute quantification of protein due to differences in ionisation efficiencies. Mass spectrometry-based proteomics, therefore, takes a comparative quantification approach using label free or label-based methods which will be covered in the following sections.

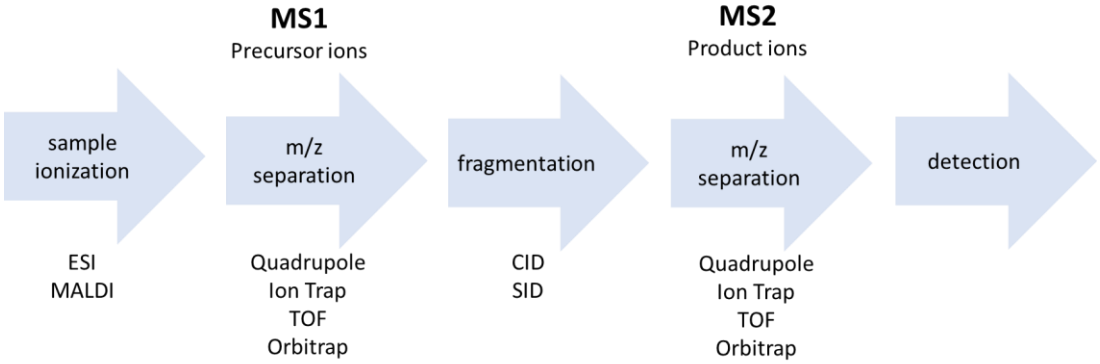


Figure 8. Steps involved in tandem mass spectrometry (MS/MS). Samples are ionised, precursor ions are separated based on mass to charge ration (m/z) in MS1, fragmented, product ions are further separated based on m/z in MS2 and detected. ESI - Electrospray ionisation, MALDI – Matrix Assisted Laser Desorption Ionisation, TOF – Time of Flight, CID – Collision Induced Dissociation, SID – Surface induced dissociation.

1.2.2.2.2.2.2 Label free quantification

In label free methods, samples are analysed independently using LC-MS/MS and protein abundances are determined by looking at the area under the curve in MS1 or spectral counting in MS2 (Figure 9). However, variation in HPLC and sample ionization is an issue in label free proteomics. Technical replicates are therefore this technique often required longer mass spectrometry run times which are expensive. Label free proteomics is however becoming increasingly popular [182].

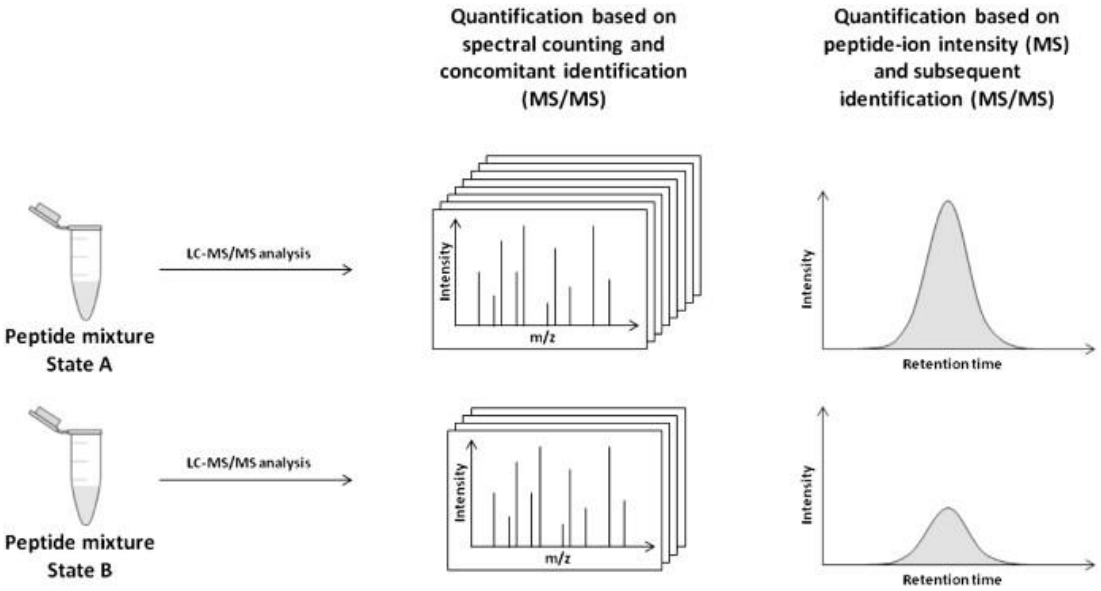


Figure 9. Label free proteomic quantification methods. Quantification can be conducted by spectral counting in MS2 or my determine area under the curve in MS1. Adapted with permission from [183]***.***

1.2.2.2.2.2.3 Label-based quantification

In label-based methods, samples are labelled, pooled, and analysed simultaneously using mass spectrometry. This removes issues of variation in mass spectra (Figure 10 Table 2), improves quantification and reduces mass spectrometry run time which can offset the additional cost of tags [184], [185]. There are two main types of tags used in proteomics: metabolic tags or tags that are chemically or enzymatically added to peptides.

One of the most common types of metabolic tags is SILAC Stable Isotope Labelling with Amino acids in Cell culture. Isotope tagged amino acids that have

a different mass to charge ratio are fed to cells/animals which allows proteins from different samples to be differentiated. Metabolically labelling samples has the advantage that it reduces issues of variability that can occur when preparing samples (Figure 10, Table 2) which means that SILAC is considered the "gold standard". However, labels are expensive, only allows for the comparison of 3 samples simultaneously and SILAC cannot be conducted on all types of samples or if cells aren't metabolically active (Table 2).

Chemical and enzymatic tags such as TMT and iTRAQ tags take a different approach and are added to peptides after digestion. These tags are isobaric and serve as reporters for protein abundance and are cleaved from peptides during MS/MS fragmentation. One main advantage of using isobaric tags is that they reduce spectral complexity. Furthermore, isobaric tags can allow for the quantification of a large number of samples as iTRAQ has been developed to tag 8 samples and TMT has recently been developed to tag 18 samples [186]. Isobaric tags do however underestimate differences in protein abundances due to precursor interference (Table 2) [184], [187]–[189]. In this study, we used iTRAQ to determine relative protein abundance which will be further discussed in the following section.

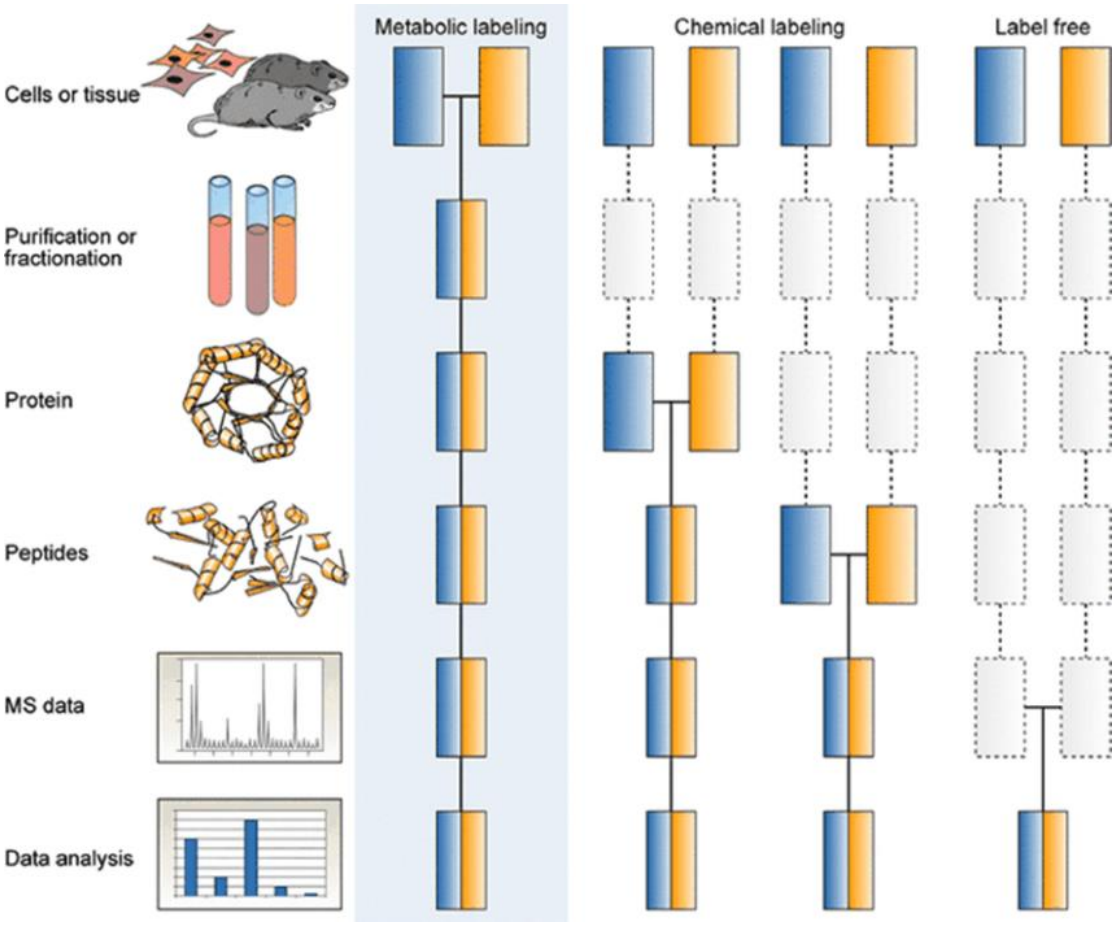


Figure 10. Mass spectrometry workflows used for relative quantification. Different samples are indicated by different coloured boxes. Blue and yellow boxes indicate when samples are combined. Sources of technical variation are indicated by dashed lines. Figure adapted with permission from [190]***.***

1.2.2.2.2.2.4 iTRAQ

iTRAQ Isobaric Tags for Relative are isobaric (have the same mass) but the location of heavy isotope varies. iTRAQ tags contain a *N*-hydroxysuccinimide (NHS) group that binds to peptides via the amine groups on the N-terminus and lysine sidechains. The same peptide from different samples that have been tagged with a unique iTRAQ tag will therefore have the same mass to charge ratio in MS1. However, upon fragmentation reporter ions of different m/z will be released allowing for relative quantification of proteins in different samples by comparing the intensity of the reporter ion peaks (Figure 5).

iTRAQ tags exist in 4-plex or 8-plex. iTRAQ reporter ions have a low mass range (m/z 113-120) to avoid interference of other ions and reporter ions of 120

m/z have been avoided due to contamination with a phenylalanine immonium ion in this area (m/z 120.08). Isotopic impurities [M+1] [M-1] of reporter ions and the phenylalanine immonium ion can interfere with reporter spectra but good resolution equipment and applying corrections overcomes this issue [189].

One important thing to consider when analysing data generated using iTRAQ and isotopic tags is that it underestimates differences in protein abundance, particularly in complex protein samples [184], [187]–[189]. This is due to precursor interference, which occurs when two or more different precursor ions are analysed simultaneously in MS2 thus affecting the intensity of reporter peaks. iTRAQ based quantification is therefore not suitable for threshold analysis as fold change values are not reliable and statistical methods should be used instead. Using mass spectrometers that further fragmenting samples in MS3 can overcome issues with precursor interference [191], [192] however, this approach was not used in this study as it requires specialist equipment that was not available to us and can reduce the number of proteins quantified [188]. None the less, regardless of this underestimation, differences in protein expression follow the same trends [184], [187]–[189] which means that iTRAQ is an ideal method to determine whether a protein is upregulated or downregulated.

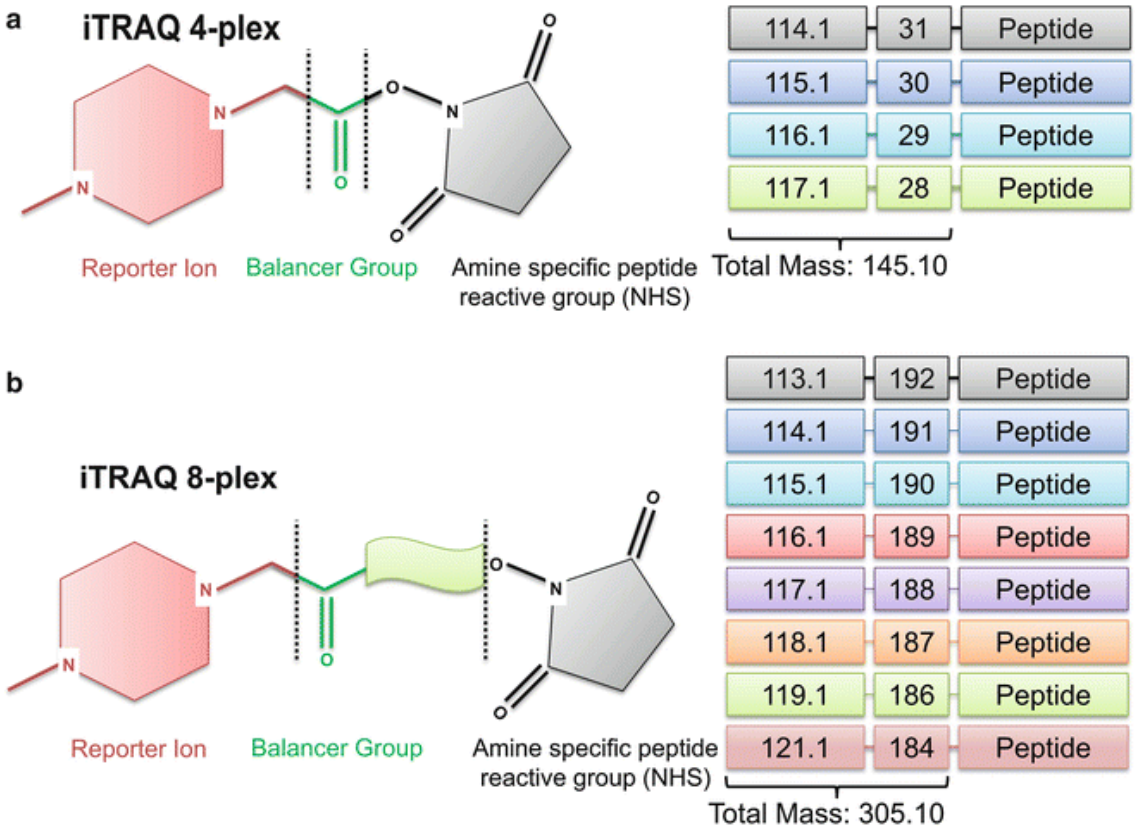


Figure 11. Structure of 4-plex (a) and 8-plex (b) iTRAQ tags. Labels are composed of a reporter ion, balancer group and NHS group that binds to amine residues in peptides. The total mass of tags is the same (145.1 for 4-plex and 305.10 for 8-plex tags) however due to differences in the positioning of heavy isotopes reporter ions have different masses (114.1-117.1 for 4-plex and 113.1-121.1 for 8-plex tags). Figure was adapted with permission from [193]***.***

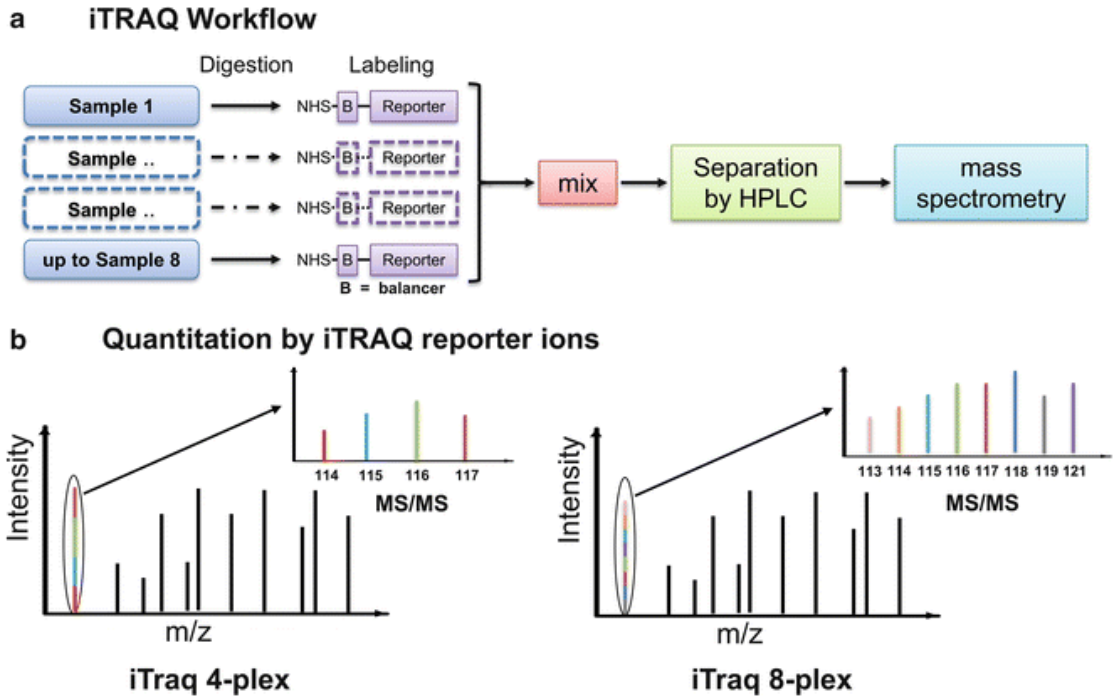


Figure 12. (a) iTRAQ workflow. (b) Reporter ions in MS2 are used to determine relative protein abundances Figure was adapted with permission from [193]***. Samples are tagged, pooled, separated using HPLC which removes any untagged peptides and analysed using mass spectrometry where the same peptide in each sample has the same mass to charge ratio (m/z) however upon fragmentation reporter ions are released which can be used to compare relative abundance of that peptide in different samples.***

1.2.2.2.3 Conclusion - Proteomics

Proteomics has provided a broader view of changes in relative protein expression which plays a critical role in cellular function. However, providing information about protein expression levels does not provide information about their activity. One factor that can affect protein activity is metabolite abundance which can be analysed using metabolomics. This will be covered in the following section.

1.2.2.3 Metabolomics

1.2.2.3.1 Metabolomics definition

Metabolomics is the analysis of multiple metabolites within a biological sample. Metabolites are small molecules that are made or used by the cell. Studying the metabolome, therefore improves our understanding of cellular physiology and enzyme activity. Unlike RNA and proteins, metabolites have

highly diverse properties. Therefore, most methods of detecting these metabolites are only able to detect specific types of molecules. Furthermore, we know very little about the metabolome and are currently only able to identify less than 20% of metabolites that are predicted in the *E. coli* metabolome database (ECMDB) [194]. Nonetheless, metabolomics provides an important insight into cellular physiology and metabolism. In the following section, we are going to discuss metabolomic sample preparation and the two main methods used in metabolomics mass spectrometry and NMR.

	Advantages	**Disadvantages**
MS	High Resolution (identify around 500 different metabolites) High dynamic range	Low reproducibility Destructive to samples More sample preparation
NMR	High reproducibility Provides structural information about metabolites Limited sample preparation Non destructive	Low sensitivity (identify around 40-200 metabolites) Low resolution Requires large sample volumes

Table 3. Advantages and disadvantages of two main methods used to analyse the metabolome: Mass spectrometry (MS) and Nuclear Magnetic Resonance (NMR).

1.2.2.3.2 Sample preparation

Metabolomic sample preparation is challenging. Certain metabolites have high turnover rates that are within the scale of seconds [195]. Samples must therefore be handled quickly and quenched to inactivate enzymes [195].

1.2.2.3.3 Mass spectrometry

Mass spectrometry can be used to identify and quantify metabolites. For a more in-depth discussion of mass spectrometry see section 1.2.2.2.2.1. Mass spectrometry techniques can identify around 500 different molecules [196] and have a high resolution allowing for the distinction of ions with a similar mass to charge ratio (for example creatine and leucine ions which have a m/z of 132.076 and 132.102 [195]) (Table 3). Furthermore, mass spectrometry techniques have a good dynamic range and are therefore able to identify low abundance proteins (Table 3) [195]–[198]. There are however some pitfalls to using mass spectrometry to identify metabolites. As mentioned previously, chromatography

and ionisation often vary between samples analysed. Furthermore, mass spectrometry is destructive to samples and provides little insight into metabolite structure (Table 3) [195]–[198].

1.2.2.3.4 Nucleic Magnetic Resonance

NMR Nuclear Magnetic Resonance is another tool used to analyse the metabolome. NMR measures how atomic nuclei respond to a radio frequency when under the influence of a magnetic field. This spectrum can then me mapped to a known spectrum to identify each molecule. The intensity of each signal can then be used to determine metabolite abundance. Unlike mass spectrometry, NMR contains more structural information which can be used to identify different metabolites and isomers [195]. NMR is also non-destructive to samples and requires less sample preparation. One of the drawbacks however is that NMR has low sensitivity making it hard to measure low abundance metabolites and large volumes of samples are therefore required (Table 3). Furthermore, NMR has limited resolution making it difficult to differentiate peaks from different molecules in complex samples and NMR studies are only able to identify 40-200 metabolites (Table 3) [196].

1.2.2.3.5 Conclusion - Metabolomics

To conclude, metabolomics is a useful tool to understand cellular physiology. However, it is only able to detect a small subset of metabolites and sample preparation is often complex. Furthermore, metabolomics on its own does not give information about metabolite flux, therefore isotope labelling is often needed to track the fate of a specific molecule. For a review on isotope labelling see [199].

1.2.2.4 Conclusion – OMICs technologies and multi-OMICs approaches

To conclude this section, different OMICs technologies (transcriptomics, proteomics, and metabolomics) present various advantages and disadvantages in studying cellular physiology. Multiple studies have therefore taken a multi-OMICs approach and combined information from different OMICs techniques,

providing a more comprehensive view of cellular physiology however integrating different types of OMICs data can be challenging [165]. The following section will discuss how OMICs technologies have been applied to improve our understanding of cellular responses to recombinant protein production in *E. coli.*

1.2.3 Using OMICs to improve recombinant protein production in *E. coli*

1.2.3.1 Overview

Multiple studies have taken OMICs approaches to examine cellular responses to recombinant protein production in *E. coli* (Table 4), (Table 5), (Table 6) and (Table 7). Most studies have used transcriptomics (Table 4) or proteomics (Table 5) to analyse cellular responses with few studies analysing the metabolome (Table 6) or taking multi-OMIC approaches (Table 7). This is probably due in part to difficulties in extracting and analysing the metabolome as well as the cost and issues with data integration associated multi-OMICs analysis.

Differences in OMICs technologies (see section 1.2.2) and variations in between conditions tested (eg. the protein overexpressed, culture conditions etc…) make it difficult to compare results from OMICs studies. Therefore the following sections will focus more on lessons learned from these OMICs studies and overarching themes.

Reference	Year	OMICs method	Recombinant protein expressed	Cellular compartment	Culture conditions	Comparison
[168]	2000	transcriptomics (microarray - 700 genes)	CAT, VP5, TMVCP, bt-FAB, IL-2	Cytoplasm, periplasm	shake flask	Soluble protein expression (IL-2, VP5, CAT) and insoluble protein expression (TMVCP), cytoplasmic expression (CAT, VP5, TMVCP, IL-2) and periplasmic expression (bt-FAB) compared to control
[169]	2000	transcriptomics (microarray- 132 genes)	LuxA	cytoplasm	shake flask	Different *E. coli* strains overexpressing LuxA compared to induced empty vector control in complex and defined media
[170]	2001	transcriptomics (microarray - 700 genes)	GFP-CAT	cytoplasm	HCDC and shake flask	Expression in high density cultures compared to shake flasks. Looked at the effect of cell conditioning (adding DTT)
[200]	2002	transcriptomics (microarray)	PLA, LCK	cytoplasm	shake flask	Soluble protein expression (PLA) compared to insoluble protein expression (LCK) and compared to non-expressing control
[201]	2003	transcriptomics (microarray)	IGF-If	cytoplasm	HCDC	compared pre induction to induced samples
[202]	2005	transcriptomics (microarray)	CAT	cytoplasm	HCDC	CAT expression compared to non-induced control.
[203]	2005	transcriptomics (microarray)	MGD-1, PepAA, pepCO	cytoplasm	shake flask	Expression of peptides with different amino acid content (PepAA least abundant amino acids, PepCO most abundant amino acids) compared to an empty vector control
[204]	2006	transcriptomics (microarray)	CAT	cytoplasm	HCDC	Comparing two methods of recombinant protein production: heat shock and dual heat shock
[205]	2007	transcriptomics (microarray)	H10257, YgfA, YjaG, YjfM, YheL, YagJ, YijU, YbjX, MSP, MSP (N36S)	cytoplasm	shake flask	Comparing cells overexpressing insoluble (YjfM, YheL, YagJ, YijU, YbjX, MSP (N36S)) and soluble (H10257, YgfA, YjaG, MSP) proteins

Reference	Year	OMICs method	Recombinant protein expressed	Cellular compartment	Culture conditions	Comparison
[206]	2009	transcriptomics (microarray)	CAT	cytoplasm	HCDC	Comparing stringent response (Serine Hydroxamate addition) to recombinant protein overexpression and a non-expressing control
[207]	2010	transcriptomics (microarray)	pTacOmpC288PL-lipase	Outer membrane (surface display)	Shake flask	Compared recombinant protein over expression to an empty vector control
[208]	2011	transcriptomics (microarray)	human interferon beta (IFNB), xylanase, GFP	cytoplasm	HCDC	Comparing high protein expression (IFNB, xylanase) to low protein expression (GFP) as well as comparing expression of proteins with different solubilities (IFNB insoluble), pre induced samples used as a control
[209]	2011	transcriptomics (microarray)	2 soluble cytoplasmic proteins 6 Inner membrane proteins	cytoplasm and inner membrane	shake flask	Comparing cells expressing different inner membrane and cytoplasmic proteins
[210]	2011	transcriptomics (microarray)	rhIFN-β	cytoplasm	HCDC	Comparing expression of the same protein with different growth rates
[211]	2020	transcriptomics (RNA-seq)	40 different proteins	cytoplasmic	shake flasks	expressing different proteins compared to empty vector control

Table 4. Transcriptomics studies observing cellular responses to recombinant protein overexpression in E. coli. HCDC – High Cell Density Cultures. Non-induced controls contain a plasmid that is not induced, empty vector controls contain an empty vector that is induced, no vector controls contain no vector and pre-induced controls are the sample before induction.

Reference	Year	OMICs method	Recombinant protein expressed	Cellular compartment	Culture conditions	Comparison
[212]	1996	proteomics 2D GE, radioactive labelling	bFGF	cytoplasm	shake flasks	Expression of bFGF compared to pre-induced and empty vector controls
[213]	2001	proteomics 2D-GE MS Pulse labelling	hFGF-2	cytoplasm	HCDC, Heat induction	Compared before and after induction (temperature upshift) and compared to an empty vector control
[148]	2001	proteomics (2D-GE MALDI MS)	F(ab')2-leucine zippper, anti-CD18, Nutropin (HGH), VEGF, IGF-1	cytoplasm/ periplasm	HCDC	Expression of different proteins compared to empty vector
[214]	2002	Proteomics (2D-GE pulse labelling)	hFGF-2	cytoplasm	HCDC, temperature induced	Expression of hFGF-2 compared to empty vector
[215]	2003	proteomics (2D-GE MS)	HGH	periplasm	HCDC	Recombinant protein production vs induced empty vector
[216]	2003	proteomics (2D-GE MS)	leptin	cytoplasm	HCDC	Expression compared to induced empty vector and pre-induced control
[217]	2003	proteomics (2D-GE)	hSOD	cytoplasmic	HCDC	hSOD overexpression over time
[146]	2005	proteomics (2D-GE)	anti-CD18F(ab')2	periplasm	HCDC	Protein overexpression over time compared to an uninduced control
[218]	2007	proteomics (2D-GE MS)	GLP-1 peptide	cytoplasm	shake flask	GLP-1 expression compared to empty vector
[219]	2007	proteomics (2D-GE MS),	YidC-GFP, YedZ-GFP, and LepI-GFP and GFP	inner membrane and cytoplasmic proteins	shake flask	Comparing cells overexpressing inner membrane (YidC-GFP, YedZ-GFP, LepI-GFP) proteins to cells expressing cytoplasmic proteins (GFP) and an empty vector control. The inclusion body proteome was also analysed
[220]	2008	proteomics (2D-GE	YidC-GFP	Membrane	Shake flask	Compared different strains (Walker and BL21(DE3)pLys)

Reference	Year	OMICs method	Recombinant protein expressed	Cellular compartment	Culture conditions	Comparison
		MS)		protein		expressing the same protein to cells not expressing protein
[221]	2010	proteomics (2D-GE MS)	GST-GlcNAc 2-epimerase-5D, GST-Neu5Ac aldolase-5R	cytoplasm	Shake flask	Comparing expression of two different proteins with different solubility to an empty vector and no vector cells
[21]	2011	proteomics (iTRAQ)	AcrA	Inner membrane	shake flask	Comparing expression of the same protein in different strains (wildtype strain and strain that glycosylated proteins) compared to an empty vector control
[222]	2013	proteomics (SILAC)	DSK2-GST	cytoplasm	shake flask	Comparing expression of a protein to pre-induced samples
[223]	2017	proteomics (2D-GE MS)	hFGF-2	cytoplasm	shake flask	Comparing expression of a recombinant protein in different media, compared to uninduced control
[90]	2018	proteomics (label free MS)	scfv-BL1	periplasm	shake flask	Comparing cells expressing high levels secreted protein (secretion system overloaded) to lower levels of expression (no secretion system overloading) and empty vector control
[44]	2019	proteomics (label free)	scfv and misfolded variant	periplasm (TAT secretion)	shake flask	Comparing cells expressing different proteins (an scFv (secreted) and an unfolded variant that remains in the cytoplasm) and an empty vector control
[107]	2019	proteomics (label free)	scfv-BL1 (ompA or dsbA signal peptide)	periplasm	shake flask	Comparing expression of the same protein with different signal peptides to an empty vector control
[106]	2020	proteomics (label free)	HGH	periplasm	24 well plate	Expression of HGH with 4 different signal peptides compared to empty vector control

Table 5. Proteomic studies observing cellular responses to recombinant protein overexpression in E. coli. HCDC – High Cell Density Cultures, 2D-GE – 2D Gel Electrophoresis, MS – Mass spectrometry. Non-induced controls contain a plasmid that is not induced, empty vector controls contain an empty vector that is induced, no vector controls contain no vector and pre-induced controls are the sample prior to induction.

Reference	Year	OMICs method	Recombinant protein expressed	Cellular compartment	Culture conditions	Comparison
[224]	2007	metabolomic (HPLC, GC–MS, isotope labelling)	hFGF-2	cytoplasm	shake flasks	Comparing hFGF-2 overexpression to pre induced and empty vector control
[225]	2011	metabolomics (GC-MS)	AcGFP1	cytoplasm	HCDC	Protein expression compared to a pre-induced control in different strains (relA knockouts) and different culture conditions over time
[226]	2011	metabolomics (MS isotope labelling)	DmpA	cytoplasmic	shake flask	Comparing expression of a protein when different amounts of inducer were added
[227]	2016	metabolomics (fourier transform infrared spectroscopy)	eGFP	cytoplasm	shake flask	The effect of riboswitches compared to no vector control
[228]	2017	metabolomics (NMR)	71 proteins	cytoplasm	shake flask	Comparing cells expressing different proteins (well-expressed and not well-expressed proteins, soluble and insoluble proteins)

Table 6. Metabolomic studies observing cellular responses to recombinant protein overexpression in E. coli. HCDC – High Cell Density Cultures, MS – Mass spectrometry. Non-induced controls contain a plasmid that is not induced, empty vector controls contain an empty vector that is induced, no vector controls contain no vector and pre-induced controls are the sample prior to induction.

Reference	Year	OMICs method	Recombinant protein expressed	Cellular compartment	Culture conditions	Comparison
[229]	2000	transcriptomics (microarray), proteomics (2D-GE MS)	α-glucosidase	cytoplasm	HCDC	Insoluble protein production over time compared to uninduced control
[230]	2008	transcriptomics (microarray), proteomics (DIGE)	hSOD and GFPmut3.1	cytoplasm	HCDC	Compared cells expressing different proteins to an empty vector induced control
[231]	2017	transcriptomics (microarray), proteomics (2D-GE MS) and	l-asparaginase	extracellular	HCDC	Comparing expression of the same protein in quiescent cells (3.5 mM indole added) to non-quiescent cells. A pre induced negative control was used
[232]	2021	Transcriptomics (microarray), proteomics (2D-GE MS pulse labelling)	human basic fibroblast growth factor	cytoplasmic	HCDC	Expression of the same protein under different fermentation conditions (carbon limited slow growth vs fast growth)

Table 7. Multi-OMIC studies observing cellular responses to recombinant protein overexpression in E. coli. HCDC – High Cell Density Cultures, 2D-GE – 2D Gel Electrophoresis, MS – Mass spectrometry. Non-induced controls contain a plasmid that is not induced, empty vector controls contain an empty vector that is induced, no vector controls contain no vector and pre-induced controls are the sample before induction.

1.2.3.2 OMICs analysis of cellular responses to recombinant protein production in *E. coli*

The most simple OMICS studies have compared *E. coli* overexpressing a model recombinant protein to a non-expressing control [106], [142], [201], [206], [207], [212], [215], [216], [218], [219], [222], [224], [227], [233]. Multiple different types of control are used in these studies including non-induced controls, empty vector controls, no vector controls, or pre-induced samples. Empty vector controls are the gold standard as cellular responses change based on plasmid presence, addition of an inducer and over time [146], [150], [155], [212], [216], [217], [225], [229]. However, these studies only provide information about *E. coli* responds to overexpression of a specific protein, under specific conditions at a specific time point.

Multiple OMICs studies have therefore compared the effect of expressing different recombinant proteins in *E. coli* [44], [107], [124], [148], [168], [200], [203], [205], [211], [221], [230]. These studies have found that cellular responses vary greatly depending on which recombinant protein is overexpressed, which could explain why methods of improving recombinant protein overexpression do not apply to all proteins. Insoluble proteins that form inclusion bodies have been shown to activate the heat shock response and based on this *ibpAB* reporters have been created to predict whether strains are overexpressing insoluble proteins [44], [168], [200], [205], [208]. Amino acid composition also affects cellular responses with there being an increase in expression of amino acid transport and synthesis and a decrease in expression of rRNA and genes involved in translation when overexpressing a peptide containing low abundance amino acids and the opposite being seen when overexpressing a peptide containing the most abundant amino acids in *E. coli* [203].

Most OMICs studies however only compare cellular responses when overexpressing a handful of proteins and it is, therefore, difficult to determine whether these responses are universal. However, one metabolomics study compared cells overexpressing 71 different proteins and used PCA analysis to identify metabolic signatures that were associated with high levels of

recombinant protein overexpression [124]. Another transcriptomics study comparing the expression of 40 different proteins, identified that increasing the expression of stress response genes was negatively correlated with successful recombinant protein overexpression [211].

Levels of protein expression have also been shown to affect cellular responses based on OMICs analysis [90], [208], [226]. Increasing the amount of recombinant protein expressed has been shown to decrease biomass formation and overall ATP concentrations based on finding from a metabolomics study [226].

OMICs studies have also identified that cellular responses vary depending on the strain used [21], [169], [220], [225]. This has also allowed us to increase our understanding of why certain strains are better at producing recombinant proteins than others beyond genomic comparisons.

Studies have also compared *E. coli* overexpressing recombinant proteins under different culture conditions [169], [170], [204], [223], [232], [234]. Culture conditions such as growth rate [232], [234], media composition [223], growth temperature [204] and growth conditions (HCDC compared to shake flasks) [170], [235] have a large effect on cellular responses with some studies observing opposite responses when using different media were used [169]. It is therefore important to consider this when designing OMICs studies.

Cellular responses change over time as OMICs analysis only provides a "snapshot" of what is happening in the cell. Therefore multiple studies have conducted OMICs analysis at different timepoints [146], [217], [225], [229]. It is important to consider at what timepoint these samples are taken (directly after induction, during growth phase or stationary phase) as this can play a large role in any changes seen. Indeed, these studies have demonstrated that cellular responses to recombinant protein overexpression change over time, the metabolome tends to vary the most over time whereas transcript and protein abundance often follow similar trends.

1.2.3.3 Cellular responses when secreting recombinant proteins

This thesis is specifically interested in improving our understanding of cellular responses to recombinant protein secretion and multiple proteomic and transcriptomic studies have covered this topic [44], [90], [106], [107], [146], [148], [168], [215]. Studies have also looked at the effect of expressing membrane proteins (inner and outer membrane proteins) and secreting recombinant proteins [21], [207], [209], [219], [220], [231] which often uses the same secretion pathways.

Multiple OMICs studies observed that secretion of recombinant proteins to the periplasm leads to an increase in expression of proteases and chaperones involved in protein folding [44], [90], [168] with certain studies indicating that secretion leads to an increase in expression of genes involved in the extracellular stress response (see section 1.1.2.1.4) in particular expression on PspA [146], [207] (for further information about the envelope stress response see section 1.1.2.1.4). Studies have also shown that recombinant protein secretion to the periplasm affects expression of genes involved in secretion with secretion of an scFv leading to an increase in expression of SecY, SecA and YidC [90] and HGH overexpression and secretion leading to an increase in expression of SecA, LepB and YidC [106]. This increase in expression of genes involved in secretion was not however seen when secreting an scFv using the TAT secretion system, leading to a mild increase in TatE and a decrease in TatAB expression [44]. Overexpression of periplasmic proteins has also been shown to lead to a decrease in expression of membrane and periplasmic proteins [44], [90], [106]. This is thought to be due to secretion system overloading preventing the export of native proteins.

OMICs studies have also improved our understanding of methods used to improve recombinant protein secretion. Reducing expression of an scFv so as to not overload the secretion system, improved yields of secreted scFv and was demonstrated to place no observable metabolic burden on the cell based on proteomics analysis [90]. It is important however to consider that proteomics only analyses a portion of the proteome and another study showed that using this method to optimize expression of a different recombinant protein (HGH) did

have an effect on the proteome, indicating that previous observations were recombinant protein specific and not widely applicable to all periplasmic recombinant proteins [106]. Multiple studies have also looked at the effect of expressing the same protein with different signal peptides to identify why adding certain signal peptides improves protein yields [106], [107]. An increase in expression of genes involved in secretion (SecA, LepB and YidC) was associated with high yields of secreted HGH when testing different signal peptides [106]. In another study, optimal secretion of an scFv when screening different signal peptides was associated with expression of genes involved in protein folding, tRNA synthesis, transcription and translation [107]. It is however difficult to determine if proteomic changes are responsible for an increase in expression of periplasmic recombinant proteins or if they are a consequence of high levels of protein expression and secretion.

However, to our knowledge, no OMICs studies have been conducted to compare expression of the same protein in the periplasm and cytoplasm to specifically understand the effect of secreting recombinant proteins on the cell. This will therefore be covered as part of this study.

1.2.3.4 Using OMICs data to direct strain engineering

OMICs data has provided a successful tool to guide genetic engineering approaches and improve recombinant protein production [201], [236] and secretion [146], [207]. These studies have improved recombinant protein overexpression by co-expressing proteins involved in amino acid synthesis such as CysK [216] and PrsA [201], nutrient uptake such as GlpF [201] and stress responses such as PspA [146], the PspABCDE and GadBC [207]. These modifications improved the rate of protein production [216] or overall protein yields, increasing yields 0.63-fold [146], 1.25-fold [201] or even 48-fold [207].

Most studies only tested the ability of their engineered strain to improve production of a specific protein [146], [201]. However, one study found that overexpressing CysK improved production of serine rich proteins (leptin and interleukin-12 β chain) but not G-CSF. Another study found that overexpressing GadBC improved activity of a lipase expressed on the outer membrane 48-fold whereas this increase in lipase activity was lower (2.51-1.59-fold) when using

different anchoring motifs [207]. It is therefore important to test whether engineered strains improve production of other recombinant proteins.

Overall, combining OMICs analysis and genetic engineering is a successful approach to improve recombinant protein production and will therefore be used as part of this study.

2 Chapter 2. Materials and methods

2.1 Strains and plasmids

2.1.1 Strains

A full list of strains used as part of this study can be found in Table 8. Further details about how knockouts were made can be found in section 2.3.11. Strains were stored at -80 °C in a 25% v/v glycerol solution in cryovials.

Strain	**Description**	**Reference**
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT*	Overexpression strain, *ompT* knockout	FUJIFILM Diosynth Biotechnologies
DH5α (*fhuA2 lac*(del)*U169 phoA glnV44 Φ80' lacZ*(del)*M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*)	Cloning and plamid storage strain	NEB
MC4100 F- [araD139]B/r DE(argF-lac)169 Lambda- e14-flhD5301 DE(fruK-yeiR)725(fruA25) relA1 rpsL150(strR) rbsR22 DE(fimB-fimE)632(::IS1) deoC1 pir^+	Storage strains for plasmids containing R6Kγ origin (pKD3, pKD4, pKD13, pKD36)	Stafford Lab
MC1000 *Δ(araA-leu)7697 [araD139]$_{B/r}$ Δ(codB-lacI)3 galK16 galE15*(GalS) *$λ^-$ e14-relA1 rpsL150*(strR) *spoT1 mcrB1*	Motile strain used as a positive control in motility assays	Stafford Lab
W3110 *λ*- IN(*rrnD-rrnE*)1 rph-1 *ΔompT ΔcpxAR*::kan	Overexpression strain, *cpxAR* knockout, kanamycin resistant	This study
W3110 λ- IN(rrnD-rrnE)1 rph-1 Δ*ompT ΔcpxP*::kan	Overexpression strain, *cpxP* knockout, kanamycin resistant	This study
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT ΔHtpX*::kan	Overexpression strain, *htpX* knockout, kanamycin resistant	This study
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT ΔpspA*::kan	Overexpression strain, *pspA* knockout, kanamycin resistant	This study
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1*	Overexpression strain, *psp*	This study

ΔompT ΔpspABCDEF::kan	operon knockout, kanamycin resistant	
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT Δtig*::kan	Overexpression strain, *tig* knockout, kanamycin resistant	This study
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT ΔyciW*::kan	Overexpression strain, *cpxAR* knockout, kanamycin resistant	This study
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT ΔyjfJ*::kan	Overexpression strain, *yjfJ* knockout, kanamycin resistant	This study
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT ΔpspA*::kan *ΔcpxAR*::cm	Overexpression strain, double *pspA* and *cpxAR* knockout, kanamycin and chloramphenicol resistant	This study
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT ΔflhDC::kan*	Overexpression strain, *flhDC* knockout, kanamycin resistant	This study
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT ΔhtpX::kan*	Overexpression strain, *htpX* knockout, kanamycin resistant	This study
W3110 *λ*- IN(*rrnD-rrnE*)1 *rph-1 ΔompT ΔyccA::kan*	Overexpression strain, *yccA* knockout, kanamycin resistant	This study

Table 8. List of E. coli strains used as part of this study.

2.1.2 Plasmids

Multiple plasmids were used as part of this study to overexpress genes and create knockouts. A full list of these plasmids can be found in Table 9.

Plasmid	Antibiotic resistance	Description	Reference
Paveway 11	TetR	Used for overexpression of model therapeutic proteins. High IPTG inducible promoter See Figure 13	FUJIFILM Diosynth Biotechnologies
Paveway 29	TetR	Used for overexpression of model therapeutic proteins. Lower IPTG inducible promoter. Same backbone as in Figure 13 but with a different promoter.	FUJIFILM Diosynth Biotechnologies
pAS15a-Cm	CmR	Backbone for co-expression of genes using constitutive promoters. See Figure 14	Based on pSB3C5-I52001
pKD46	AmpR	Helper plasmid for lambda red recombineering, temperature sensitive	[237]
pKD4, pKD13	KmR, AmpR	Plasmids used to amplify constructs to make kanamycin resistant knockouts	[237]
pKD3, pKD32	CmR, AmpR	Plasmids used to amplify constructs to make chloramphenicol resistant knockouts	[237]

Table 9. List of plasmids and their antibiotic resistance used in this study.

Expression plasmids (Paveway 11 and Paveway 29) from FUJIFILM Diosynth Biotechnologies were used to overexpress and secrete model therapeutic proteins (eg. scFv, HGH, sfGFP and Hel4) (Figure 13) (see Appendix 1 for the sequence of these model proteins). Paveway plasmids have a high copy origin of replication and contain a tetracycline resistance marker. Paveway 11 has a high IPTG inducible promoter and Paveway 29 has a lower IPTG inducible promoter. Paveway 11 was used for most of this study.

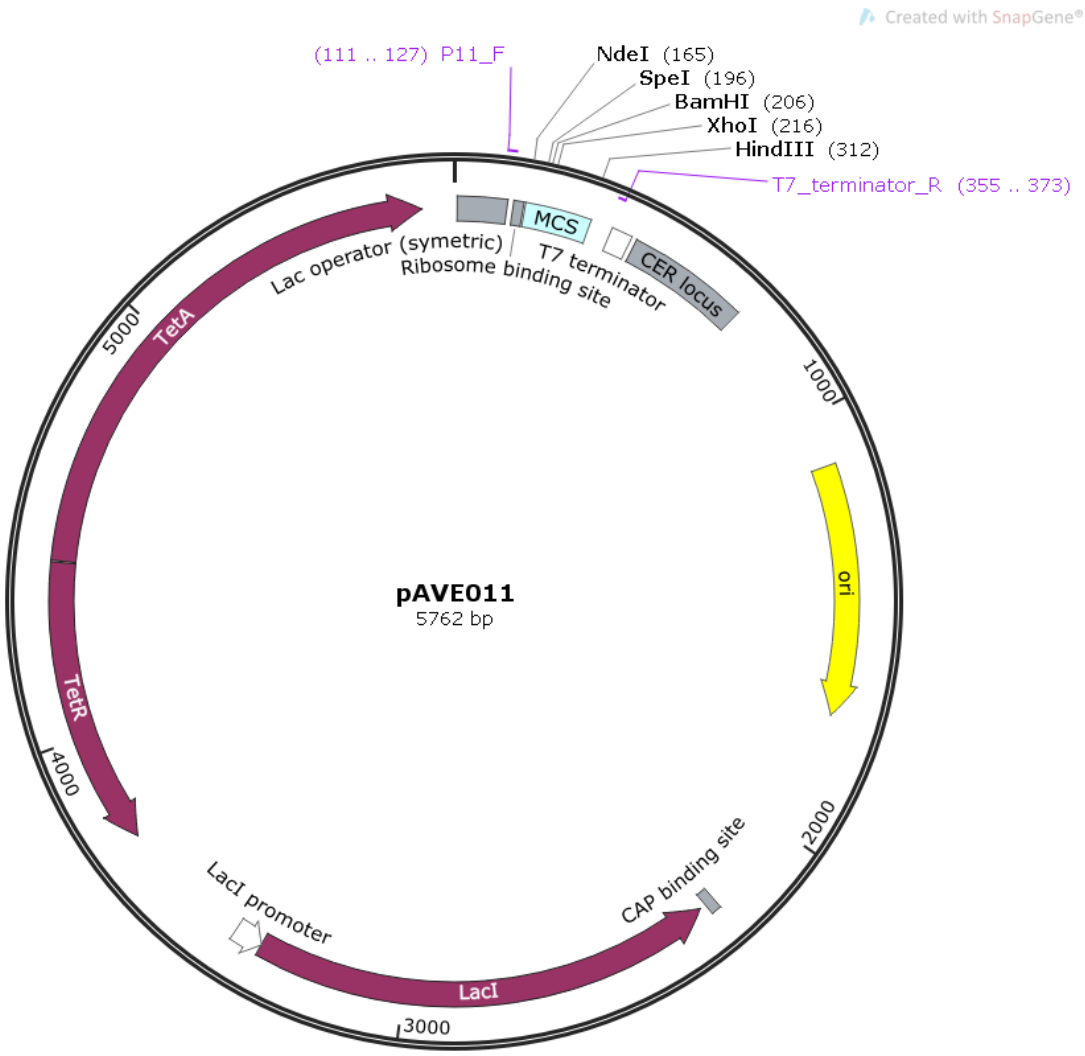


Figure 13. Plasmid map of Paveway 11 and its multiple cloning site (MCS). This plasmid map was constructed using SnapGene. Ori indicates the origin of replication, TetA and TetR confer resistance to tetracycline, LacI is the lactose repressor P11_F and T7_terminator_R are primer binding sites.

A plasmid was also designed to co-express genes from *E. coli* (pAS15a-Cm) using constitutive promoters of various strengths (Anderson promoters) [238]. pAS15a-Cm is compatible with the paveway 11 plasmid as it contains a p15a low copy origin of replication and a chloramphenicol resistance marker. This plasmid was based on the iGEM pSB3C5-I52001 plasmid, but the CcdB selection marker and high copy origin of replication were removed to reduce gene synthesis costs. Additional restriction sites were also added to facilitate cloning.

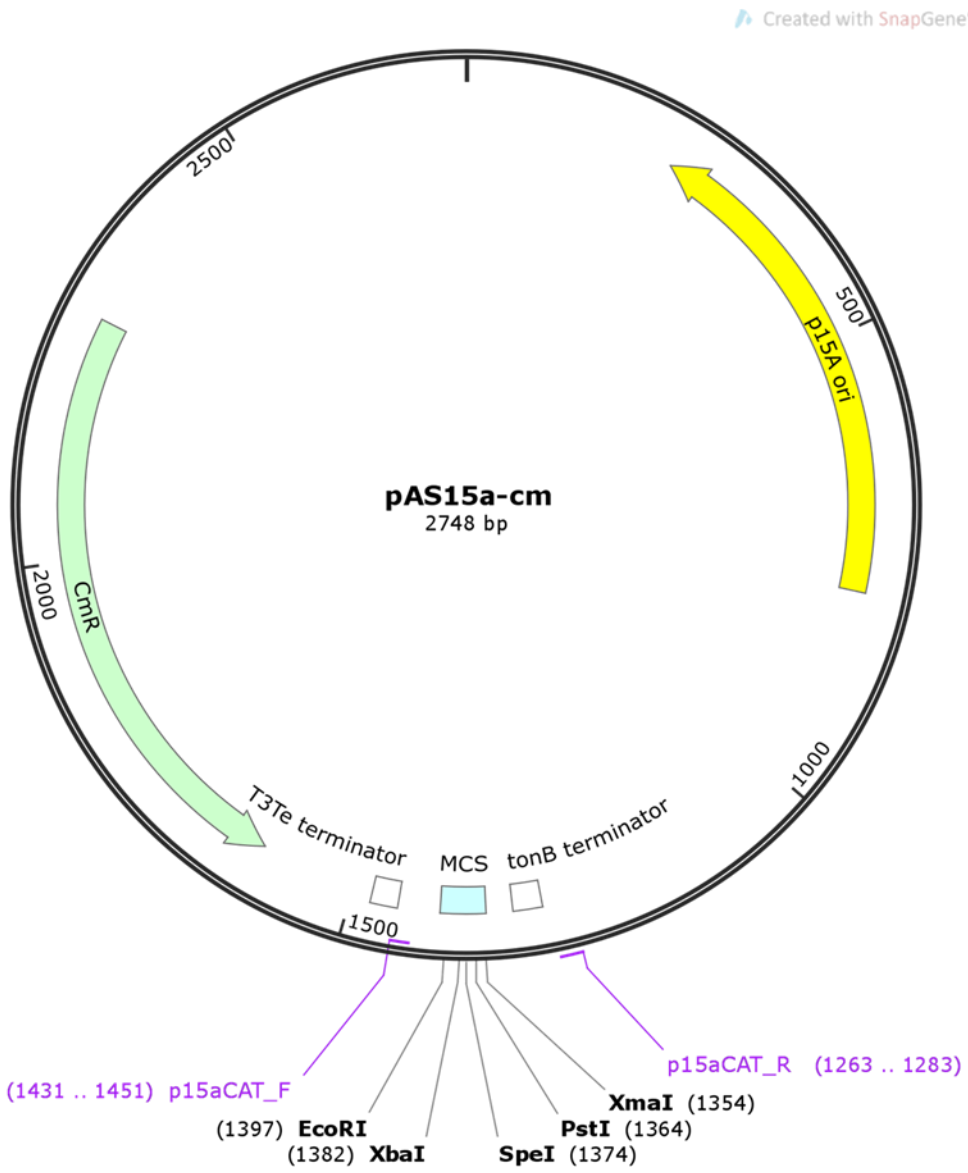


Figure 14. Plasmid map of pAS15a-Cm and its multiple cloning site (MCS). This plasmid map was constructed using SnapGene. Ori indicates the p15a low copy origin of replication, CmR indicates the chloramphenicol resistance gene, MCS is the multiple cloning site and pAS15a_F and pAS15a_R are primer binding sites.

2.2 Media and culture conditions

2.2.1 Media

Bacteria were routinely grown in LB broth (Miller) which was made by resuspending 25 g of LB powder (Fisher) in 1 L deionised water. Strains were also streaked out on LB agar plates which was made by resuspending 40 g LB agar powder (Merck – formerly Sigma) in 1 L deionised water.

To determine optimal culture conditions to test strains, bacteria were grown in media similar to that used in HCDC that contained glycerol, yeast extract and M9 salts (Table 10). 100mM solution of sodium hydroxide, 6-6.6% solution of ammonium hydroxide and 10x MOPS buffer (83.72 g MOPS, 7.17 g Tricine, 450 mL dH_2O adjusted to pH 7.4 with 10M KOH) was added to media to test if this improved bacterial growth.

Component	**Amount per L**
Glycerol	20 g
Disodium phosphate	6.8 g
Monopotassium phosphate	3 g
Ammonium chloride	1 g
Sodium chloride	0.5 g
Magnesium Sulphate	0.240 g
BD bacto yeast extract 212750	5 g

Table 10. Composition of media used to optimize growth conditions in shake flasks.

Similar media was used in Ambr fermentations at FUJIFILM Diosynth Biotechnologies (20 g/L BD Bacto yeast extract, 35 g/L glycerol, M9 salts, trace metals, antifoam, $CaCl_2$ and $MgSO_2$.) however specific media composition cannot be discussed here. All media except the trace metals were autoclaved to sterilise. This medium was also used to test alterations to strains and media in shake flasks and 96 well plates as it was more representative of Ambr conditions, however trace metals and antifoam were not added as these either did not work under my hands or were not available.

2.2.1 Antibiotic stock solutions

Antibiotics were added to media to select for specific plasmids and mutations. A list of the antibiotics and the concentrations used can be found in Table 11. Antibiotic stocks were filter sterilised using a 0.22 μm syringe filter (Millipore).

Antibiotic	Stock solution	Storage temperature	Concentration used
Ampicillin	50 mg/mL in water	-20 °C	200 µg/mL
Chloramphenicol	50 mg/mL in 100 % ethanol	-20 °C	25 µg/mL
Kanamycin	50 mg/mL in water	4 °C	50 µg/mL
Tetracycline	12 mg/mL in 70 % ethanol	-20 °C	10 µg/mL

Table 11. List of antibiotic stock solution and concentrations used.

2.2.3 Culture conditions – HCDC – Industrially relevant conditions

Growth of high cell density cultures was conducted in Ambr250 (Satorius) fed batch bioreactors at Fujifilm Diosynth Biotechnologies following the protocol described in patent WO2007/088371 with some modifications to media composition which cannot be discussed here. Media contained 20 g/L BD Bacto yeast extract, 35 g/L glycerol, M9 salts, trace metals, antifoam, $CaCl_2$ and $MgSO_2$. 750 µL of overnight cultures were inoculated into 150 µL of media and samples were grown at 37 °C. The total dissolved oxygen level (pO_2) was kept above 30% through oxygen supplementation, stirring and addition of 50% v/v Mazu DF 204 antifoam in ethanol. The pH was maintained at 6.7 by addition of 14% w/w ammonium hydroxide. Upon depletion (OD600 of 50 +/-5), expression was induced with 0.125 mM IPTG or 0.0125 mM IPTG (125 mM IPTG stock solution – filter sterilised). Upon induction, the temperature was reduced to 30 °C and a glycerol and ammonium sulphate solution was gradually added to promote growth. 5 mL samples were taken every 12 hours post induction for a total of 48 hours.

2.2.4 Culture conditions – Shake flask – Testing engineered strains and media

Shake flask conditions were used to test engineered strains and media for improved therapeutic protein production and secretion to the periplasm.

Bacteria were grown in 250 mL baffled flasks (Merck - formerly Sigma, manufactures code 1134/08) as shown in Figure 15. Shake flask conditions were initially optimized using 50 mL of media described in (Table 10). Subsequently strains were tested in 20 mL of media that was the same as that used in Ambr HCDC (an M9 based media supplemented with yeast extract and glycerol) minus the trace metals and antifoam as trace metals did not solubilize under my hands and Mazu antifoam was not available. Flasks were inoculated with an overnight culture to make a bacterial density (OD_{600}) of 0.05 and grown at 37 °C in a shaking incubator set to 150 rpm (SciQuip Incu-Sshake TL6-5). During mid log phase at around 3-4 OD_{600} nm, samples were induced with 0.125 mM or 0.0125 mM IPTG (100 mM stock solution – filter sterilised). The temperature in the fermenters was then reduced to 30 °C. Samples were taken pre-induction, 3 hours post-induction and 24 hours post-induction where the optical density at 600 nm was measured see section 2.2.2 and samples were pelleted at 11 000 xg for 5 mins. Samples were then further processed as described is section 2.4.

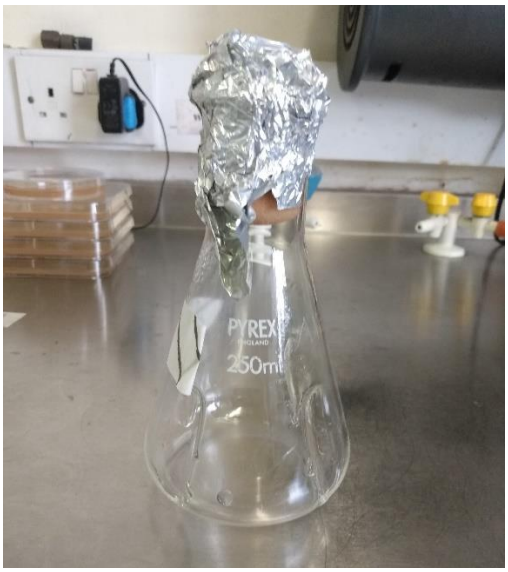


Figure 15. 250 mL baffled shake flask used in this study.

2.2.1 Culture conditions – 96 well plate – Growth curves

Growth curves were conducted in a 96 well plates using a Tecan Sunrise plate reader. The same media was used as in the fermenters (an M9 based media supplemented with yeast extract and glycerol) minus the trace metals and

antifoam. Overnight cultures were diluted to the equivalent of an OD_{600} nm of 0.05. Wells at the perimeter of the plate were filled with 200 µL of media to avoid drying of samples (Figure 16). 200 µL of diluted bacterial cultures were added to each well in triplicate and all other wells were filled with 200 µL of media. The plate was covered with a projector sheet that had been cut to size to avoid drying of samples, allow for aeration and avoid condensation. The projector sheet was sterilised by spraying with 70% IMS and placing on a UV box for 10 mins and was attached to the plate with tape at the top and sides as seen in Figure 16. Bacteria were grown at 37 °C shaking using the "normal" shaking speed on the Sunrise plate reader. OD_{600} readings were taken every 30 mins for a total of 20 hours.

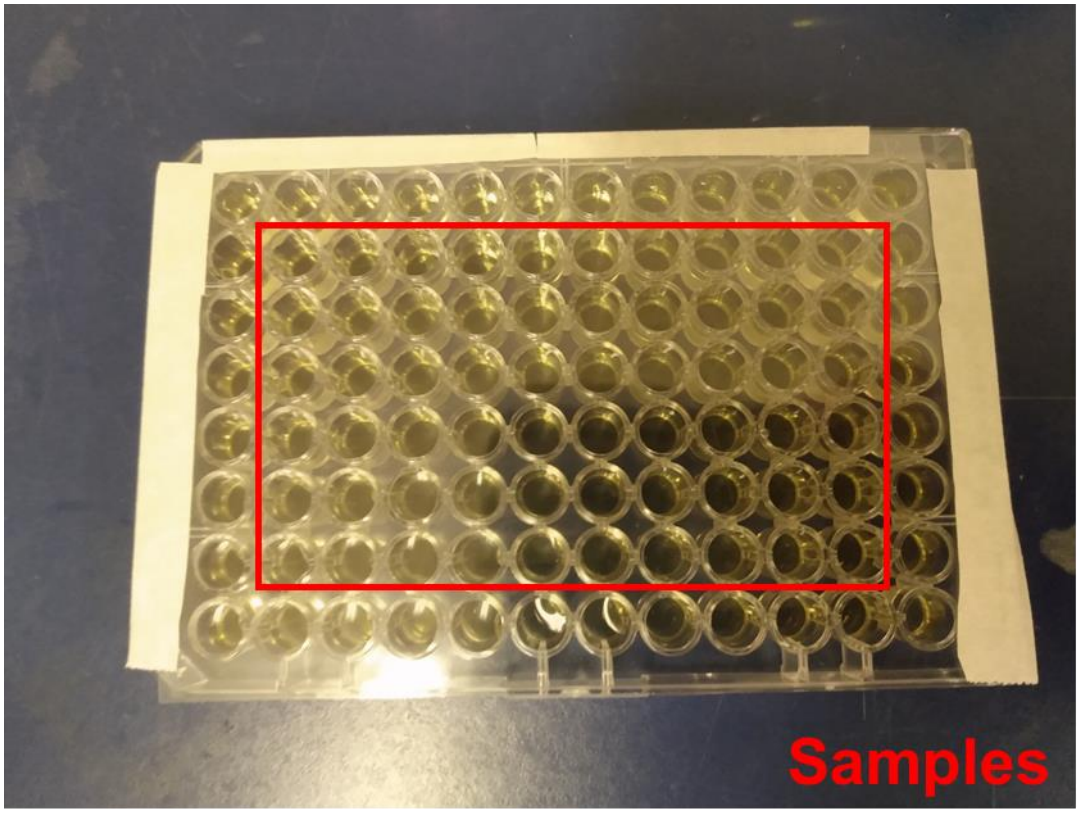


Figure 16. Setup of 96 well plate used for growth curves. The red box surrounds wells containing samples.

2.2.2 Determining bacterial concentrations – OD_{600} and dry cell weights

Bacterial concentrations were determined using the optical density at 600 nm (OD_{600}). Optical densities above 1 were further diluted to remain within the accurate range (OD 0.1-1).

Dry cell weights were determined at FUJIFILM Diosynth Biotechnologies by measuring the weight of an empty Eppendorf, pelleting 1 mL of bacterial at maximum speed for 5 mins, removing the supernatant, and drying pellets at 105 °C for over 48 hours. The dry cell weight was then calculated by subtracting the weight of the tube from the weight of the tube containing the dry pellet.

2.2.3 Measuring supernatant pH using a colour indicator

Supernatant pH in shake flasks was measured using a colour indicator. 1 mL of *E. coli* cultures was centrifuged at 11 000 xg for 1 min. The supernatant was then mixed with 200 µL of Fisher full range indicator (U/0025/PB08) and pH was determined based on colour change.

2.3 Techniques used for cloning

2.3.1 Genomic DNA extraction

Genomic DNA was extracted from *E. coli* to use as a template for PCR. Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega) following the standard protocol for gram negative bacteria. DNA quality and quantity were determined using nanodrop following the standard protocol for nucleic acid. DNA was considered pure if it had a 260/230 value between 2.0-2.2 and 260/280 value between 1.8 and 2.0.

2.3.2 Plasmid extraction

Plasmids were extracted with Isolate II plasmid mini kit (Bioline) using the standard protocol. The low copy protocol was used for low copy plasmids

(eg. pAS15a-Cm, pKD46). Plasmid quantity and quality were determined using nanodrop and/or by agarose gel electrophoresis (see section 2.3.6).

2.3.3 Polymerase chain reaction - PCR

Polymerase chain reaction (PCR) was used to amplify specific genes from genomic DNA or plasmids.

Primers were supplied from Merck (formerly Sigma) and were designed to have a GC content of 40-60% and annealing temperatures (Tm) between 45-72 °C with a maximum 4 °C difference in annealing temperature between primer pairs. Primers were also checked for any major secondary structures and dimers using the IDT OligoAnalyser tool and ran through the NCBI BLASTn tool to check their specificity. A list of primers used to amplify sequences to clone into plasmids can be found in Table 12.

Name	**Sequence**
scFv_F	ggaattccatatggatattcagatgacccaatcc
scFv_R	ccgctcgagttattagcgcttgatctcca
scFv_FLAG_R	ccgctcgagttaatggtggtggtgatgatggcgcttgatctccacc
pspA_F	gctctagatgggtatttttttctcgctttgcc
pspA_R	tgcactgcagttattgattgtcttgcttcattttgg

Table 12. List of primers used for cloning genes into plasmids. Restriction sites are underlined, and additional nucleotides were added to the end of each primer to improve digestion.

Phusion high fidelity polymerase (NEB) was used to amplify constructs with high fidelity following the standard protocol. A master mix was created using the protocol found in Table 13 with 50 ng - 250 ng of genomic DNA or 1 pg - 10 ng of plasmid DNA added as a template. Negative controls containing no template were also used. Samples were then placed in a thermocycler following the standard cycle found in Table 14. The specific annealing temperature used for each primer pair was calculated using the NEB annealing temperature (Tm) Calculator. Samples were then separated using agarose gels to check for amplification of products of the correct size. If the PCR was unsuccessful, a

gradient PCR with annealing temperatures within a +/- 5 °C range of the predicted Tm was used to determine the optimal annealing temperature.

Component	50 µL reaction
Nuclease free H_2O	Made up to 50 µL
5x phusion HF buffer	10 µL
10 mM dNTP	1 µL
10 µM Forward primer	2.5 µL
10 µM Reverse primer	2.5 µL
Template	variable
Phusion polymerase*	0.5 µL

*Always added at the end

Table 13. Phusion PCR master mix.

Step	Temperature	Duration	Number of cycles
Initial denaturation	98 °C	30 seconds	1
Denaturation	98 °C	10 seconds	25-35
Annealing	45-72 °C	30 seconds	
Extension	72 °C	30 seconds per 1 kb	
Final extension	72 °C	10 minutes	1
Hold	5-10 °C	∞	1

Table 14. Thermocycling conditions for Phusion PCR.

2.3.4 Annealing oligos

Oligos were annealed to create small sections of double stranded DNA (<100 bp). For each sequence, an oligo as well as its reverse complement was synthesised by Merck (formerly Sigma). A full list of these primers and their reverse complement can be found in Table 15. Oligos were resuspended in 1x Buffer 4 (NEB) and an equimolar amount of each oligo were mixed in an eppendorf. Samples were then placed on a heat block at 95 °C for 5 minutes,

removed from the heat block and left to cool to room temperature. Samples were then separated on a 2% agarose gel (see section 2.3.6) and a band of the correct size was cut out and extracted from the gel (see section 2.3.7).

Name	Sequence
High_promoter_BBa_J23119	ggaattcttgacagctagctcagtcctaggtataatgctagcttctagagaaagaggggacaaactagtc
Medium_promoter_BBa_J23106	ggaattctttacggctagctcagtcctaggtatagtgctagcttctagagaaagaggggacaaactagtc
Low_promoter_BBa_J23103	ggaattcctgatagctagctcagtcctagggattatgctagcttctagagaaagaggggacaaactagtc
High_promoter_BBa_J23119_RC	gactagtttgtcccctctttctctagaagctagcattatacctaggactgagctagctgtcaagaattcc
Medium_promoter_BBa_J23106_RC	gactagtttgtcccctctttctctagaagctagcactatacctaggactgagctagccgtaaagaattcc
Low_promoter_BBa_J23103_RC	gactagtttgtcccctctttctctagaagctagcataatccctaggactgagctagctatcaggaattcc

Table 15. List of oligos and their reverse complement (RC) that were annealed to make double stranded DNA. Restriction sites are underlined.

2.3.5 Gene synthesis

Genes were synthesised and cloned into the appropriate vector by Genewiz. A list of synthesised genes can be found in Appendix 2.

2.3.6 Agarose gel electrophoresis

Agarose gels were used to separate DNA and RNA. 1% w/v agarose gels were used large constructs (more than 500 bp) and 2% w/v agarose gels were used for smaller constructs (less than 500 bp). The appropriate weight of agarose was dissolved in 1x TAE (see Table 16) by heating in the microwave. 1 µL of ethidium bromide was added per 100 mL of gel and poured into a gel tray to set. Once set, the gel was placed into a gel box filled with 1x TAE (see Table 16). Samples were then resuspended in 5x loading dye (except for DreamTaq PCR samples) and 10-50 µL of the sample was added to each well with 5 µL of Generuller 1kb ladder (ThermoFisher) or PCRBio IV (PCR biosystems)

depending on the expected size of the product. The gels were then run at 80 V for 1 hour before being imaged under UV.

Component	Amount per L
Tris base	4.84 g
Glacial acetic acid (17.4 M)	1.14 mL
EDTA disodium salt	0.37 g

Table 16 1xTAE composition.

2.3.7 DNA clean-up

PCR products were cleaned up before restriction digestion or sequencing using the isolate II PCR and gel Kit (Bioline). DNA was directly isolated for the PCR reaction when only one band was present on the agarose gel following the standard protocol. However, when multiple products were present on the gel, the band of the correct size was cut out of the gel using a scalpel and isolated with the same kit following the protocol for gel extraction.

2.3.8 Restriction digestion

All restriction enzymes used were supplied by NEB. Restriction enzymes were chosen that had unique restriction sites to the areas being cloned (see multiple cloning sites in plasmid maps Figure 13 and Figure 14). Double restriction digestion was conducted using the standard NEB protocol found in Table 8 and using the appropriate NEBuffer for the restriction enzymes that were used. This reaction was then incubated for 1 hour at 37 °C. 12.5 units of Antarctic Phosphatase (NEB) were then added to 50 µL of digested plasmid to avoid relegation of plasmid and incubated for 30 mins at 37 °C. A DNA clean-up kit was then used to remove restriction enzymes from the reaction (see section 2.3.7).

Component	50 µL reaction
Nuclease free H_2O	Make up to 50 µL
10X NEBuffer	5 µL
DNA	1 µg
Restriction enzyme	1 µL per restriction enzyme used

Table 17. Restriction digestion reaction.

2.3.9 DNA ligation

DNA was ligated using a T7 DNA ligase kit (NEB) using the standard protocol. The ligation reaction was made following the protocol found in Table 18 with as much of the DNA insert added to the reaction as possible. Samples were then incubated for 15-30 mins at room temperature (25 °C).

Component	20 µL reaction
2X T7 DNA ligase buffer*	10 µL
Vector DNA	50 ng
Insert DNA	made up to 20 µL
T7 DNA ligase**	1 µL

*aliquoted to avoid freeze thawing as it contains ATP

**added last

Table 18. T7 DNA ligation reaction.

2.3.10 Transformation

Ligated plasmids were transformed into 5-alpha Competent *E. coli* (High Efficiency) cells (NEB) following the standard protocol. Cells were incubated with 5 µL of ligation mix on ice for 30 mins before being heat shocked at 42 °C for 30 seconds and incubated on ice for 2 minutes. Bacteria were then resuspended in 1 mL LB or SOB and grown at 37 °C for 1 hour. The cells were then pelleted at 11 000 xg for 1 minute, the supernatant was removed, and the pellet was resuspended in 100 µL of broth. This was then plated on the appropriate antibiotic selection plate and incubated overnight at 37 °C (see section Table 11)

All other constructs were transformed into chemically competent cells or electrocompetent cells that were made using the following methods.

2.3.10.1 Chemically competent cells and heat shock

Chemically competent cells were prepared following a standard protocol. LB broth was inoculated with an overnight culture of *E. coli* and grown to mid log phase OD_{600} (0.6-0.8). Cells were kept on ice for 15 mins, pelleted by centrifugation at 3000 xg 5 mins at 4 °C, washed 3 times with an equal volume of ice cold 0.1 M $CaCl_2$ and concentrated 100-fold by resuspending in ice cold 0.1 M $CaCl_2$ 20% v/v glycerol. Chemically competent cells were aliquoted into 50 µL aliquots and stored at -80 °C. Thawed competent cells were incubated with up to 1-5 µL of plasmid on ice for 30 mins, heat shocked at 42 °C for 30 seconds and incubated on ice for 5 seconds. Bacteria were then resuspended in 1 mL LB or SOB and grown at 37 °C or 30 °C for 1 hour 100 µL of bacterial culture was incubated overnight at 37 °C or 30 °C on the appropriate antibiotic selection plate (see section Table 11).

2.3.10.2 Electrocompetent cell and electroporation

Electrocompetent cells were made using a standard protocol where LB broth was inoculated with an overnight culture of *E. coli* and grown till mid log phase OD_{600} (0.6-0.8). Cells were kept on ice for 15 mins, pelleted by centrifuging at 3000 xg 15 mins at 4 °C, washed 3 times with an equal volume of ice-cold deionised water and concentrated 250-fold by resuspending in ice cold 20% glycerol. Electrocompetent cells were aliquoted into 50 µL aliquots and stored at -80 °C. 50 µL of electrocompetent cells were mixed with 2 µL of plasmid in a chilled tube. Samples were then added to 0.2 mm prechilled electroporation cuvettes. Cells were then electroporated at 1.8 kV for more than 5 seconds. 1 mL LB was immediately added to the cuvette and cells were grown shaking at 37 °C or 30 °C for 1 hour before 100 µL of bacterial culture was incubated overnight at 37 °C or 30 °C on the appropriate antibiotic selection plate (see section Table 11).

2.3.11 Lambda red recombineering

Lambda red recombineering was conducted following the standard protocol from [239]. Knockouts were created using PCR to amplify antibiotic resistance genes (chloramphenicol or kanamycin) flanked by regions that were homologous to the area upstream and downstream of the gene being knocked out. This was done by designing primers that contained 36-50 bp homologous to the area directly upstream or downstream of the gene being knocked out as well as priming sites (priming site 1, 2 and 4) that were used to amplify antibiotic resistance genes from pKD3, pKD4, pKD13 or pKD32. pKD3 and pKD32 were used to create chloramphenicol resistant knockouts and pKD4 and pKD13 were used to create kanamycin resistance knockouts. pKD3 and pKD4 were specifically used to knockout genes that were part of an operon as they contained a ribosome binding site and start codon that would allow transcription of downstream genes.

A full list of primers used to create knockouts can be found in Table 19. Amplification of these PCR products was conducted using Phusion high fidelity PCR (see section 2.3.3). PCR products were then run on a 1% agarose gel to check proper amplification and bands were cut out of the gel and extracted to avoid contamination of samples with the template (see sections 2.3.3 and 2.3.7).

Name	**Sequence**
*pspA*_knockout_F	tcaaaaagataaaaaattggcacgcaaattgtattaacagttc**gtgtaggctggagctgcttc**
*pspA*_knockout_R	aaaacaaacagcacaaaaatggttaacggaatagccagaaatagcgcgct**catatgaatatcctccttag**
*psp*_operon_knockout_F	ccgaataaagcattcacgccgcatccggcaagttgtattgctcaacttcg**gtgtaggctggagctgcttc**
*psp*_operon_knockout_R	gtgtgacagaaaaaaaaacggcgcataagcgccgctcatggtgaattctt**attccggggatccgtcgacc**
*tig*_knockout_F	gcttgcggggtaagagttgaccgagcactgtgatttttgaggtaacaag**gtgtaggctggagctgcttc**
*tig*_knockout_R	cctggcggtgacgggcctttgtgcgaatttagcgcgttatgctgcgtaaa**attccgggatccgtcgacc**
*cpxP*_knockout_F	aatcgtaaactctctatcgttgaatcgcgacagaaagattttgggagcaa**gtgtaggctggagctgcttc**
*cpxP*_knockout_R	catgtgggggaagacagggatggtgtctatggcaaggaaaacagggttta**catatgaatatcctccttag**
*yjfJ*_knockout_F	tggctgaaattacagaagaatattcacactaattacaaggacgggtattt**gtgtaggctggagctgcttc**
*yjfJ*_knockout_R	cagacgctggaaaaaaccagacatattttatccttaaaatagttgccagc**attccgggatccgtcgacc**
*yciW*_knockout_F	ttcccgccttgcgtcaggataacgatttcctttacgaccaaggagcgccc**gtgtaggctggagctgcttc**
*yciW*_knockout_R	gccattagagcggctgacagttttacgcgaatctgtctgacgcggcaagg**attccgggatccgtcgacc**
*cpxAR*_knockout_F	agtcatggattagcgacgtctgatgacgtaatttctgcctcggaggtatt**gtgtaggctggagctgcttc**
*cpxAR*_knockout_R	ggcgtaaacgccttatcctgcctgcaaatgcgaagt**attccggggatccgtcgacc**
*htpX*_knockout_F	cttgaaaatagtcgcgtaacccatacgatgtgggtatcgca**gtgtaggctggagctgcttc**
*htpX*_knockout_R	tgtcagtggtttacgtatcactcagccacgatccagggcataaaaaaagc**attccgggatccgtcgacc**
*yccA*_knockout_F	attatcattatgctgcttaattaattacatctgtcatagagagtgactca**gtgtaggctggagctgcttc**
*yccA*_knockout_R	ctaaggaggatattcatatgtctcacccgctaacactttcatagcctcgctttatgcgg**attccggggatccgtcgacc**

*flhDC*_knockout_F	ggtgcggtgaaaccgcataaaaataaagttggttattctgggtgggaata**gtgtag gctggagctgcttc**
flhDC_knockout_R	gcagcggtacgtcgttaccgctgctggaatgttgcgcctcaccgtatcag**attccg gggatccgtcgacc**

Table 19. List of primers used to create knockouts. Priming sites are highlighted in bold.

The recombineering plasmid pKD46, was transformed into cells using either heat shock or electroporation (see section 2.3.10) and grown on ampicillin plates at 30 °C. Successful transformants were grown in LB-amp at 30 °C overnight. These overnight cultures were then inoculated into fresh LB-amp media containing 10 mM arabinose (1M arabinose stock solution- filter sterilised) to induce expression of lambda red genes. Cultures were then grown at 30 °C in a shaking incubator till cultures reached an OD_{600} 0.6-0.8. Cells were then made electrocompetent and transformed with 10-100 ng of purified PCR product as per described in section 2.3.10.2. Once electroporated 1 mL LB was added to the cells and these were grown in a shaking incubator for one hour at 37 °C. Cells were then plated out onto the appropriate selection plate (kanamycin and/or chloramphenicol) and grown at 37 °C overnight to remove the temperature sensitive pKD46 plasmid. Colonies were then streaked out onto kanamycin or chloramphenicol selection plates as well as ampicillin plates to check whether the pKD46 ampicillin resistance plasmid remained in the cells. Colonies that retained ampicillin resistance, were grown at 42 °C and re-streaked on ampicillin plates to check for retention of the pKD46 plasmid. Knockouts were then further tested using colony PCR.

2.3.12 Colony PCR

Colony PCR was used to check cloning and creation of knockouts. When checking cloning, primers were designed that flanked the multiple cloning site (Table 20). When checking knockouts, primers were designed that flanked the gene that had been knocked as well as primers within antibiotic resistance cassette (k1, k2, c1, c2) in case the gene that was knocked out was a similar size to the antibiotic resistance cassette (Table 21).

Name	Sequence
Paveway_11_F	gctcacaattccccact
T7_terminator_R	gctagttattgctcagcgg
pAS15a_F	gccagtgagttgattgctacg
pAS15a_R	agccccaatgataaccccaag

Table 20. List of primers used to verify multiple cloning sites.

Name	Sequence
PspF_up_F	catccggcaagttgtattgctc
PspA_down_R	gccagcgagtattcataactttcctcac
PspE_down_R	gagatcggtttgcgggtcaac
Tig_up_F	gcgaagccaacaacctg
Tig_down_R	ggaaccgtgcgaaaagc
CpxP_up_F	cgtaaactctctatcgttgaatcg
CpxP_down_R	gcttccgttatactagcgtcagt
YjfJ_up_F	aacctcgctggtagataacg
YjfJ_down_R	ctaacgtgtcgagcgtaaagc
YciW_up_F	cgccgatccgtctttctcc
YciW_down_R	gcagcgggttgtcctgaaac
CpxAR_up_F	ggctgcaaacatgcgtcagg
CpxAR_down_R	tcaggcatcctgctcaaatgc
HtpX_up_F	ggtgacttacgcactatccagac
HtpX_down_R	tgcctgtgtcagtggtttacg
YccA_up_F	ttctgggtgcccttacg
YccA_down_R	caatcaccacagccagc
FlhDC_up_F	gcagcttatcgcaactattctaatgc
FlhDC_down_R	gacaggatgttcagtcgtcagg
k1	cagtcatagccgaatagcct
k2	cggtgccctgaatgaactgc
c1	ttatacgcaaggcgacaagg
c2	gatcttccgtcacaggtagg

Table 21. List of primers used to check knockouts.

Colony PCR was conducted using 2x DreamTaq master mix (ThermoFisher Scientific) following the standard protocol. A master mix was created using the protocol in Table 22. Individual colonies were then added to the reaction using a toothpick. When conducting colony PCR to check knockouts, colonies were resuspended in 50 µL nuclease free water and heated for 10 mins at 95 °C to improve cell lysis which was added to the master mix instead of the nuclease free water.

The standard thermocycler cycle found in Table 23 was used with the appropriate annealing temperature for the primers used. Samples were then analysed using agarose gels to check for amplification of products of the correct size. PCR products were then purified following the standard protocol found in section 2.3.7 and sanger sequencing (Eurofins Genomics) of PCR products with both the forward and reverse primer was used to further confirm clones and knockouts.

Component	20 µL reaction
Nuclease free H_2O	8 µL
10 µM Forward primer	1 µL
10 µM Reverse primer	1 µL
2x DreamTaq master mix	10 µL

Table 22. DreamTaq PCR Master mix.

Step	Temperature	Duration	Number of cycles
Initial denaturation	95 °C	3 minuites	1
Denaturation	95 °C	30 seconds	25-40
Annealing	45-72 °C	30 seconds	
Extension	72 °C	1 minute up to 2 kb	
Final extension	72 °C	15 minutes	1
Hold	5-10 °C	∞	1

Table 23. Thermocycling conditions for DreamTaq PCR.

2.4 Protein extraction and analysis

2.4.1 Protein extraction

2.4.1.1 Whole cell and supernatant

Whole cell and extracellular proteins were prepared by centrifuging cells at 16 000 xg for 5 mins. The supernatant was then decanted into a tube and both the supernatant and the cell pellet were stored at -20 °C.

2.4.1.2 Periplasmic extraction and spheroplasts

Periplasmic proteins and spheroplasts were extracted from fresh bacterial cultures using cold osmotic shock. The equivalent of 1 mL of bacteria at OD_{600} 60 was pelleted at 14 000 xg for 5 mins and the supernatant was discarded. Osmotic shock solution 1 (OS1) was made following the protocol described in Table 24, sterile filtered and stored at 4 °C. The pellet was gently resuspended in 1 mL osmotic shock solution 1 and incubated for 20-27 hours at 4 °C. Cells were then pelleted at 14 000 xg for 5 mins and the supernatant was aliquoted into tubes. The pellet was then gently resuspended in 1 mL osmotic shock solution 2 (OS2 1.23 g/L $MgSO_4.7H_2O$), incubated on ice for 10 mins and then centrifuged for 5 mins at 14 000 xg. OS1 and OS2 fractions were then mixed in equal volumes and stored at -20 °C. The remaining pellet containing the spheroplast was also stored at -20 °C.

Component	**Amount added per L**
Trizma-Base	12.14 g
Trizma-HCl	1.57 g
Tetra-sodium EDTA di-hydrate	0.41 g
Sucrose	200 g

Table 24. Osmotic shock solution 1 (OS1).

2.4.1.3 Soluble and insoluble protein extraction

Soluble and insoluble proteins were extracted by using chemical lysis to disrupt cells and high-speed centrifugation to pellet insoluble proteins. Cells

were chemically lysed using BugBuster protein extraction reagent (Norvagen). 1 tablet of cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Merck) was added to 10 mL of BugBuster reagent (Norvagen). Cells were resuspended in 5 mL of BugBuster reaction mix per gram wet cell weight of pellet (lysozyme was not added as this has a similar molecular weight to the scFv). Cells were then incubated on a rotating platform for 10 mins and unlysed cells were pelleted by centrifuging at 3 000 xg for 5 mins. 100 µL of the supernatant was added to a new tube and centrifuged at 16 000 xg for 20 mins at 4 °C. The supernatant containing soluble proteins was then decanted and stored at -20 °C. The pellet was then resuspended 10 mls of 1:10 diluted BugBuster per gram of wet cell pellet and centrifuged at 16 000 xg for 15 mins at 4 °C. This step was further repeated to wash the pellet 3 times. The supernatant was removed, and the pellet was stored at -20 °C.

2.4.2 Protein analysis

2.4.2.1 SDS-PAGE

SDS-PAGE was used to separate protein samples. Samples were run on either 12% or 4-20% mini-PROTEAN TGX precast gels (Bio-Rad) and hand cast 12% or 18% tris/glycine gels depending on the size of the protein analysed. To create hand cast gels, a resolving gel was made following the protocol in Table 25 and isopropanol was added on top of this so that samples ran in a straight line. The stacking gel was made following the protocol in Table 25 pipetted on the resolving gel and a comb was placed on top. 2x SDS loading dye (see recipe in Table 26) was added to each sample and heated at 95 °C for 10 mins. A list of the concentrations of each sample used can be found in Table 27. 5 µL of prime step protein latter (Biolegend) and 10 µL sample were added to each well. Gels were run at 150V in a tank containing running buffer (3.02g Tris base, 18.8g glycine and 1 g SDS made up to 1 L with water) for an hour or until proteins ran to the bottom of the gel. SDS-PAGE gels were stained with 5 mL of Instant Blue (Expedion) for 1 hour before being washed with deionised water and imaged.

Component	12% resolving gel	18% resolving gel
40% Acrylamide/bis solution	3 mL	3.75 mL
Lower tris buffer (181.8 g/L Tris base, 4g/L SDS pH8.8)	2.5 mL	2.5 mL
H_2O	4.5	3.55 mL
TEMED*	5 µL	5 µL
10% w/v Ammonium persulphate *	350 µL	350 µL

Component	Stacking gel
40% Acrylamide/bis solution	0.975 mL
Upper tris buffer (60.6 g/L Tris base, 4 g/L SDS pH 6.8)	2.1 mL
H_2O	4.725 mL
TEMED*	17 µL
10% w/v Ammonium persulphate *	100 µL

*Added at the end to polymerise gels, 10% Ammonium persulphate solution was made fresh

Table 25. Composition of resolving and stacking gel. Makes two 1 mm thick gels.

Component	Amount added
1M DTT	1 mL
20% SDS	1 mL
glycerol	2 mL
0.5 M Tris HCL pH 6.8	1.25 mL
0.02 % bromophenol blue	200 µL
cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail	1 tablet
Nuclease free water	4.55 mL

Table 26. 2x SDS lysis buffer

Sample	Sample dilution - SDS-page gel
supernatant	Normalised to OD_{600} of 40 and suspended in 2x SDS lysis buffer
pellet	The equivalent of 1 mL of culture at an OD_{600} of 0.1 was pelleted and resuspended in 1 mL of SDS lysis buffer (Ambr samples) or the equivalent of 1 mL of culture at OD_{600} of 1 was suspended in 200 µL of SDS lysis buffer (overexpression strains) and 10 µL was added to each well.
periplasm	Re-suspended in 5x SDS lysis buffer (National Diagnostics)
spheroplast	The equivalent of 1 mL of culture diluted to an OD_{600} of 0.1 was pelleted and resuspended in 1 mL of 2x SDS lysis buffer
insoluble	200 µL cell lysate was pellet resuspended in 40 µL 2x SDS lysis buffer
soluble	Diluted 1/20 in 2x SDS lysis buffer

Table 27. Sample concentrations used for SDS-PAGE analysis.

2.4.2.2 Western blots

Western blots were conducted to detect specific proteins. Acrylamide gels were made and run following the standard protocol for SDS-PAGE found in section 2.4.2.1. Unstained gels were placed on top of a 0.45 µm Amersham nitrocellulose membrane and sandwiched between 6 layers of filter paper soaked in semi dry transfer buffer (recipe found in Table 28). The proteins were then transferred to the membrane using the Bio-rad transfer system and ran on a semidry transfer cycle (2.5 A 10 V 20 mins). To check transfer and equal loading of lanes, Ponceau stain (ThermoFisher) was added to the membrane and washed off using deionised water. The membrane was then incubated rocking for one hour at room temperature or overnight at 4 °C in 20 mL 5% milk powder TBST (1 x TBS (see recipe Table 29) 0.1% v/v TWEEN 20). The membrane was then incubated for 1 hour on a rocker in a 1:10 000 dilution of anti-PspA antibody (supplied by Prof. Martin Buck), 1:80 000 dilution of anti-GroEL antibody (Merck - formerly Sigma G6532-.5ML), 1:1000 anti-FLAG HRP conjugated (sigma A8592), 1:5000 anti-his tag (3D5) HRP conjugated (ThermoFisher R931-25), 1:1000 anti-igkv1-5 (Biorbyt, orb184430) 20 µg anti-HGH antibody (MAB1067 RnDSystems) in 10 mL in 5% milk powder TBST. The membrane was then washed three times for 5 mins in TBST. The membrane was then incubated for 30 mins rocking with a 1:3000 dilution of anti-rabbit HRP antibody (Cell

Signalling 7074) (PspA, GroEL blots) or 1:3000 anti-mouse HRP antibody (Cell Signalling 7076P2) (HGH blots) in 10 mL 5% milk powder TBST. The membrane was then washed two times for 5 mins in TBST and then for 5 mins in TBS. The secondary antibody step was skipped for the anti-his tag antibody as this was already HRP conjugated. The ladder was then drawn onto the membrane using a ECL pen (Licor West Sure pen). Most blots were incubated for 5 mins in 2 mL Pierce ECL western blotting substrate (ThermoFisher) following the standard protocol and anti-PspA blots were incubated with a more sensitive Westar Supernova ECL substrate (Cyanagen). Blots were imaged on a c-digit using the long run (12 mins).

Component	Amount added
TRIS Base	5.8 g
Glycine	2.9 g
SDS	0.37 g
methanol	200 mL
H_2O	Made up to 1 L

Table 28. Semi-dry transfer buffer.

Component	Amount added
Tris	60.5 g
NaCl	87.6 g
H_2O*	Made up to 1 L

Table 29. 10 x TBS adjusted to pH 7.9

2.4.2.3 Labchip capillary electrophoresis

LabChip capillary electrophoresis (Perkin Elmer) was conducted at FUJIFILM Diosynth Biotechnologies using the standard protocol to separate proteins. Spheroplast and whole cells were resuspended in water and the bacterial concentration was adjusted to an OD_{600} of 1, whereas supernatants and periplasmic fraction were not diluted. 1.5 µL of the sample was added to 52.5 µL of buffer T (24.5 µL of DTT added to 700 µL of protein express samples buffer) and incubated at 100 °C for 5 mins. 18 µL of this mixture was then added to 70 °C

water and samples were then analysed using LabChip capillary electrophoresis with frequent washes between samples.

2.4.2.4 Image analysis of band intensity

ImageJ was used to measure intensity of scFv bands on SDS-PAGE gels using densitometry analysis. Images were converted to greyscale (image>type>8-bit) and background intensity was removed using the standard rolling ball method in ImageJ (process>subtract background). Each lane was selected using the same sized rectangle and average intensities were plotted using the gels function in the analyse tab. A vertical line was drawn between scFv high and low bands and whole cell proteins and the area under the curve was measured to determine intensity of the high scFv band (contains an OmpA signal peptide 14 kDa), the low scFv band (no signal peptide 12 kDa) and whole cell proteins (excluding the 14 kDa and 12 kDa scFv) (Figure 17). The intensity of the secreted scFv band (12 kDa) was normalised to the intensity of whole cell proteins and the intensity of the secreted scFv band (12 kDa) was also normalised to the intensity of the non-secreted scFv band (14 kDa) to reduce any variation due to differences in protein loading. Some lanes were however difficult to analyse due to gel streakiness affecting average intensity calculations or curved bands affecting resolution of peaks and were therefore excluded from analysis.

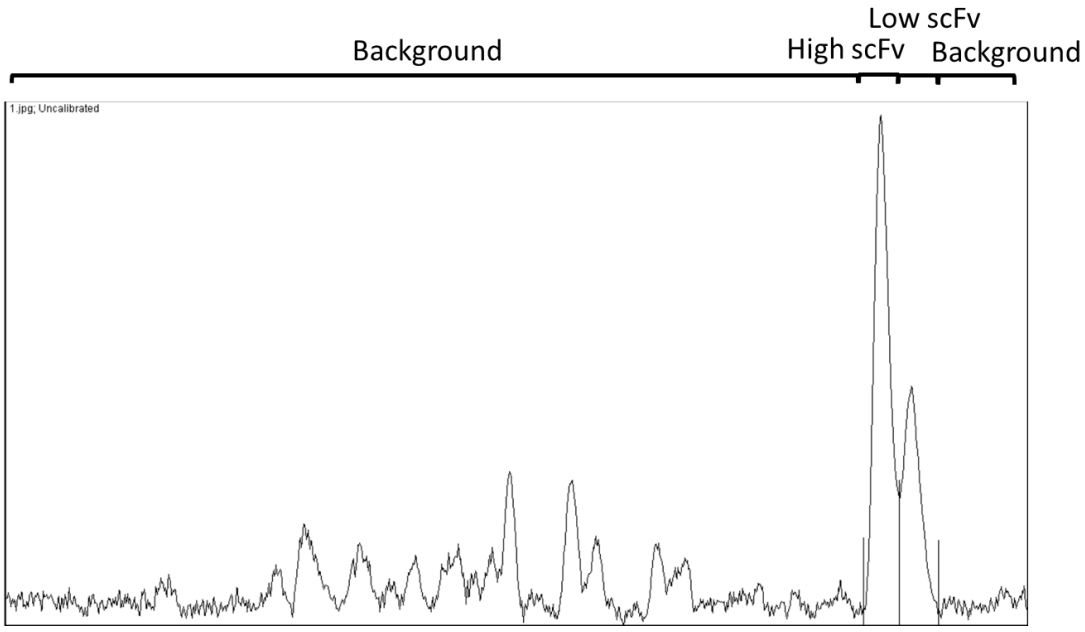


Figure 17. Image analysis of band intensity differentiating whole cell proteins, high scFv (14 kDa-contains the OmpA signal peptide) and low scFv bands (12 kDa).

2.5 Transmission electron microscopy - TEM

Unless otherwise specified all reagents used for TEM were supplied by Agar Scientific. Bacteria were grown in shake flask following the standard protocol (see section 2.2.4). Samples were induced with 0.125 or 0.0125 mM IPTG and the equivalent of 1 mL of bacteria at 1 OD_{600} was pelleted 3 hours post-induction by centrifuging at 6 000 xg for 5 mins. The supernatant was discarded, and pellets were fixed and cacodylate buffer (0.1 M Cacodylate buffer, pH7.4) and (2.5 % glutaraldehyde) for a minimum of 4 hours at 4 °C. Pellets were then washed 3 times with cacodylate buffer. The buffer was removed, and pellets were stained using osmium tetraoxide (OsO_4) for 1:30 hour – 2 hours at room temperature. Pellets were then dried by incubating for 15 minutes in increasing concentrations of ethanol (50 %, 75%, 95% and 100% ethanol). Samples were incubated in propylene oxide for 15 min at room temperature and then further incubated overnight in 50/50 reagent mixture of propylene oxide/resin in sealed containers. Resin was always made fresh and was made by mixing 50/50 v/v mixture of DDSA and aldehyde resin with one drop of accelerator (BDMA) per mL of resin. The following day the propylene oxide/resin mixture was removed, and the pellets were incubated twice in fresh resin for 3 hours. Pellets were

placed in fresh resin in a flat embedding silicone mould (Agar Scientific) and were set by incubating at 60°C for 48 h.

Blocks were then trimmed and polished using a scalpel and a microtome (Reichert-Jurg Ulracut E) fitted with a glass knife. Samples were then sectioned to a 90 nm thickness on the microtome using a diamond knife (DiATOME). Sections were expanded using vapour from a cotton bud dipped in chloroform and placed on formvar-coated grids.

Grids were stained by incubating them face down on a drop of uranyl acetate (7% saturated in 50% methanol) for 30 mins in the dark before being washed by placing them on a drop of deionised water for 30 minutes. Grids were then placed on a drop of lead citrate (see Table 30) for 5 mins in the dark before being washed for 5 minutes in deionised water. Between each step excess stain and water were blotted from the grids using filter paper.

Samples were then imaged using a Tecnai T12 Spirit transmission electron microscope.

Component	**Amount added**
Lead nitrate	1.33 g
Sodium citrate	1.76 g
Distilled H_2O	30 mL
1 M Sodium Hydroxide (NaOH)	Make up to 50 mL

Table 30. Reynolds lead citrate

2.6 Motility assay

Bacteria were spotted onto 0.4% w/v agar plates and incubated at 30 °C overnight. Zones of motility were measured using a ruler. *E. coli* MC1000 strains were used as a positive control for motility.

2.7 RT-qPCR

2.7.1 RNA extraction

Cells were stored using the RNAprotect Bacterial reagent (Qiagen) using the standard protocol. 2 volumes of RNAprotect were added to 1 volume of

bacterial culture and vortexed for 5 seconds before being incubated at room temperature for 5 mins. Samples were then pelleted by centrifugation at 5000 xg for 5 minutes before the supernatant was removed and pellets were stored at -80 °C.

RNA was extracted using the RNeasy Kit (Qiagen). The equivalent of 1 mL of bacteria at an OD_{600} of 0.5 was used for each extraction. Cells were lysed using lysozyme (Lysozyme from chicken egg white – Merck (formerly Sigma)) and proteinase K (Qiagen) following the standard protocol. On column DNase digestion was conducted using the RNase-Free DNase set (Qiagen) to remove DNA contaminants. If DNA contamination was still present, further DNase digestion was conducted in solution using the same RNase kit and samples were purified using the RNeasy columns.

Sample quantity and quality were analysed using Nanodrop and acrylamide gels. RNA samples were prepared by adding 2x RNA loading dye (ThermoFisher) to each sample and heating for 5 min at 65 °C before running on a 1% agarose gel (see section 2.3.6 – no ethidium bromide was added). Nanodrop™ was conducted following the standard protocol (see section 2.3.1) and RNA with an OD260/280 around 2 was considered pure.

2.7.2 cDNA synthesis

cDNA synthesis was conducted using the standard protocol for the Applied Biosystems High-Capacity cDNA RT kit. A master mix was made following the protocol as seen in Table 31. 10 µL of the master mix was then mixed with 10 µL of sample (100 ng of RNA). A control containing no reverse transcriptase was conducted for all RNA samples to check for DNA contamination. A water control was also used. Samples were then amplified in a thermocycler following the protocol found in Table 32. cDNA samples were then stored at -18 °C.

Component	Amount added
10x RT buffer	2 µL
100 mM dNTP Mix	0.8 µL
10x RT random primers	2 µL
MultiScribe reverse transcriptase	1 µL
Nuclease free H_2O	4.2 µL

Table 31. cDNA synthesis master mix

Cycle temperature	Duration
25 °C	10 mins
37 °C	2 hours
85 °C	5 mins
4 °C	∞

Table 32. Thermal cycling conditions cDNA synthesis.

2.7.3 qPCR- sybr green

qPCR was conducted using the PCR biosystems qPCRBIO SyGreen Mix Lo-ROX following the standard protocol. A master mix was created following the protocol found in Table 33 and 1 µL of cDNA was added to each 10 µL reaction. All measurements were conducted in triplicate. Primers were designed to have an annealing temperature within the range of 59-61 °C, amplify an 80-150 bp product and were unique to the gene of interest based on blast analysis. A list of these primers can be found in Table 34 with RpoA serving as an internal control. Water controls were also used for each primer pair to check for contamination of the master mix. No reverse transcriptase controls were conducted for each cDNA sample using the RpoA primer set to check for DNA contamination. qPCR was conducted using a RotorGene-Q thermal cycler with a 2-step protocol and a melt curve which can be found in Table 35. The threshold was set mid exponential phase. Certain samples were also separated on a 2 % DNA gel (see section 2.3.6) to check that the correct product was amplified. Ct values were determined by setting the threshold mid exponential phase. Samples were considered contaminated if the water and no

transcriptase control had a Ct value lower than 27. 2^ΔCt calculations were used to determine transcript abundance.

Component	Ammount added
2x qPCRBIO SyGreen Mix	5 µL
Forward primer (10 µM)	0.4 µL
Reverse primer (10 µM)	0.4 µL
cDNA	1 µL
Nuclease free H_2O	3.2 µL

Table 33. Sybr green master mix for qPCR.

Name	Sequence
RpoA_F	gtcgacgcatgctacagc
RpoA_R	atttcgatgaccagcttgtcc
PspA_F	gacatcgtgaatgccaac
PspA_R	caaccagtgtatcttccatctc

Table 34. List of primers used for qPCR.

Step	Temperature	Duration	Number of cycles
Polymerase activation	95 °C	2 mins	1
Denaturation	95 °C	5 sec	40
Anneal/Extension	59-61 °C	30 sec	
Melt curve			1

Table 35. 2-step qPCR protocol with melt curve.

2.8 Sample preparation transcriptomics

2.8.1 RNA extraction

RNA was extracted using the RNeasy Kit (Qiagen) following the protocol found in section 2.7.1. When preparing RNA from Ambr bioreactor samples 4 volumes of RNA protect were added per volume of media. Some samples were stored in RNA protect for a maximum of 7 hours (12 and 36 hours post-induction samples) as these samples were taken overnight.

RNA quantity and quality were analysed using Nanodrop (ThermoFisher) with a OD260/280 around 2 and 260/230 value between 2.0-2.2 being considered pure.

RNA quality and quantity of RNA-seq samples were also analysed using Qubit (ThermoFisher scientific) and labchip capillary electrophoresis (PerkinElmer) by Genewiz following their standard protocols.

2.8.2 RNA sequencing

Further sample preparation and RNA-sequencing was conducted by Genewiz. RNA depletion was achieved using Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat) (Illumina). Library preparation was conducted using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB). Samples were fragmented and underwent cDNA synthesis. Samples were then adenylated, and universal adapters were ligated to the cDNA. RNA sequencing was conducted using Illumina HiSeq 4000 in paired end mode.

2.8.3 Data analysis

Most data analysis was also conducted by Genewiz. Sample quality was analysed using FastQC and MultiQC [240], [241]. Reads were trimmed to remove adapters and nucleotides with poor quality using Trimmomatic v.0.36 [242]. Reads were mapped to the *Escherichia coli* W3110 genome (ensemble bacteria version 47) [243] as well as the scFv sequence with and without the signal peptide using STAR aligner v.2.5.2b [244]. Unique gene hit counts were

calculated using Subread package v.1.5.2 [245] FeatureCounts function. Differential gene expression analysis was conducted using the R package DESeq2 [246]. Samples were not deemed differentially expressed if their adjusted pvalue ≥ 0.05 and Log2 fold change was between 1 and -1. Heatmaps, PCA and Volcano plots were generated using the following tools: the heatmap.2() function in the gplots package [247] the plotPCA() function of the DESeq2 [246] and the EnhancedVolcano R package [248].

2.9 Sample preparation for proteomics

2.9.1 Protein extraction – bead beating

Cells pellets were resuspended in 1 mM tri-ethyl ammonium bicarbonate buffer (TEAB) pH 8.5 ± 0.1, containing 0.05% SDS and diluted to make 500 µL with an OD_{600} of 20 in low bind Eppendorfs (Eppendorf). 0.7 g glass beads 200-300 µm (Merck, formerly Sigma Aldrich) were then added to the tubes and bead beating was conducted using a Scientific Industries™ Genie Disruptor for 45 seconds with 3 minutes incubation on ice between each round of bead beating, 10 times. Samples were then sonicated in an ice-cold ultrasonic bath 3 times for 3 minutes with 2 minutes of incubation on ice between each sonication. Samples were then centrifuged at 21 000 xg for 15 minutes at 4 °C to remove unlysed cells and insoluble proteins. The supernatant was collected in 1.5 mL low bind Eppendorfs and stored at -80 °C. The protein concentration of each sample was determined using a Bradford assay following the standard protocol. A standard curve was created 0 to 2.1 µg BSA in 0.3 µg increments. Samples were also diluted 1:10, 1:20, 1:40 and 1:100. 10 µL of sample or standard was mixed with 200 µL of protein assay reagent (Bio-rad), in triplicate. Samples were then incubated for 5 mins and absorbance at 595 nm was measured using a plate reader. 10 µg of sample was then separated on an SDS-PAGE (see section 2.4.2.1) to check for equal loading.

2.9.2 Reduction, alkylation and tagging of samples

55 µg of each sample were adjusted to the same volume using 1 mM TEAB pH 8.5 ± 0.1, containing 0.05% SDS. Samples were reduced by adding $1/10^{th}$ of

the samples volume of 50 mM TCEP and incubating at 60 °C for 1 hour. Samples were then alkylated by adding 1/20th of the sample volume of 200 mM MMTS and incubated for 20 mins at room temperature. 2.75 grams of Pierce trypsin protease (MS grade) (Thermo Fisher Scientific) was added to each sample to make up 1/20 of the protein weight and samples were digested by incubating overnight at 37 °C. Samples were then separated on an SDS-PAGE following the standard protocol in section 2.4.2.1 to check for protein degradation. If bands still appeared on the SDS-PAGE gel after digestion a further 1/100 of trypsin was added to the samples and incubated for 3 hours. Samples were tagged with 8-plex iTRAQ tags (ABSciex). 50 µL of isopropanol was added to each iTRAQ tag and the contents of each tube were added to each sample, vortexed and incubated at room temperature for 2 hours. Samples were then pooled and vortexed to mix.

2.9.3 Sample fractionation – reverse phase HPLC.

Pooled samples were fractionated based on hydrophobicity using a Hypercarb™ Porous Graphitic Carbon LC reversed phase Analytical Column (Cat no. 35003-052130, ThermoFisher Scientific, UK), with 3 µm particle size, 50 mm length, 2.1 mm diameter and 250 Å pore size. Fractionation was conducted using a Dionex UltiMate 3000 Autosampler linked to Dionex UltiMate 3000 Flow Manager and Pump system (Thermo Scientific), which was washed through with 20% CAN. Samples were re-suspended in 60 µL Buffer A (3% acetonitrile, 0.1% Trifluoroacetic acid) and 57 µL of the sample was loaded onto the column.

Buffer A was exchanged with Buffer B (97% acetonitrile, 0.1% Trifluoroacetic acid) with a flow rate of 0.2 mL min^{-1} with the following gradient: 2% B (0-15 minutes), 2-30% B (15-80 minutes), 30-60% B (80-130 minutes), 60-90% B (130-131 minutes), 90% B (131-136 minutes), 2% B (137-145 minutes).

Fractions were collected every two minutes from 20 minutes to 120 minutes and dried at room temperature in an Eppendorf vacuum centrifuge. 5 µL Loading Buffer (3% acetonitrile, 0.1% Trifluoroacetic acid) was then added to each sample. 2 µL of each fraction were then pooled and combined into 9

samples (F1-F9) so that samples represented a broad range of retention times and peaks intensities (214 nm and 280 nm) F1 (42, 58, 102, 104 mins), F2 (46, 60, 74, 88 mins), F3 (48, 62, 76, 90 mins), F4 (50, 64, 78, 92 mins), F5 (52, 68, 80, 94 mins), F6 (56, 68, 82, 96 mins), F7 (54, 70, 84, 98 mins), F8 (44, 72, 86, 100 mins) and F9 (106, 108, 110, 112 mins).

2.9.4 LC-MS/MS

LC-MS/MS was conducted using nano-flow liquid chromatography (U3000 RSLCnano, Thermo Scientific). 2 µL of each sample was loaded onto an Easy-Spray C18 column (75 µm x 50 cm) and ran on a 2-step gradient (97% solvent A (0.1% formic acid in water) to 10% solvent B (0.08% formic acid in 80% acetronitrile) over 5 min and then 10% to 50% B over 75 min at a flow rate of 300 nL min^{-1}. The chromatography column was coupled to a quadrupole-orbitrap mass spectrometer (Q Exactive HF, Thermo Scientific) that was programmed for data dependent acquisition with 10 product ion scans (resolution 15000, automatic gain control 5e4, maximum injection time 20 ms, isolation window 1.2 Th, normalised collision energy 32, intensity threshold 2.5e5) per full MS scan (resolution 60000, automatic gain control 3e6, maximum injection time 20 or 100 ms). Proteins were then identified against the *E. coli* W3110 genome and scFv sequence with and without the signal peptide using MaxQuant [249]. Perseus was used to remove contaminants and duplicates from data as well as to conduct hierarchical clustering and PCA. Differentially expressed genes were then determined using a Linear Models For Microarray Data (LIMMA) approach using ProteoSign following the standard protocol [250].

2.10 Functional OMICs analysis

Differentially expressed genes were further analysed using Ecocyc (omics dashboard (Gene ontology), regulatory overview and cellular overview (metabolic map)) [251]–[253], TFinfer to determine overrepresented regulatory pathways [254] and ClusterProfiler to identify enriched Gene Ontology (GO) and KEGG terms [255]

2.11 Statistical analysis

Statistical analysis was conducted to compare growth of different strains in different culture media and compare SDS-PAGE and western blot band intensity. Statistical analysis was conducted using GraphPad Prism. A basic T-test was used to compare two samples, one-way ANOVA was used to compare multiple samples and two-way ANOVA was used to compare grouped samples in growth curves. Differences were determined as statistically different when $p \leq 0.05$ and are indicated by a star in graphs.

3 Chapter 3 Multi-OMICs analysis of cellular responses to recombinant protein secretion to the periplasm

3.1 Introduction

Protein overexpression is known to affect the cellular physiology of *E. coli*, activating stress responses [157] and placing a metabolic burden on the cell [150], [156]. There has therefore been a growing interest to improve our understanding of cellular responses to recombinant protein overexpression to improve yields of therapeutic proteins [160], [161]. OMICs tools such as transcriptomics, proteomics and metabolomics have drastically increased our understanding of how *E. coli* responds to recombinant protein production and how this response varies depending on the protein that is overexpressed and growth conditions (see section 1.2.3 of the introduction).

Many therapeutic proteins are secreted to the periplasm to improve yields, ease purification and produce proteins containing disulphide bonds such as antibody fragments which make up >35% of therapeutic proteins produced in *E. coli* [1]. Overexpression of proteins in the periplasm can lead to multiple issues such as periplasmic inclusion body formation [66], [93]–[95] and overloading of the secretion system [63], [89], [90], [97], [98] (see section 1.1.3 of the introduction). Multiple studies have used OMICs to compare *E. coli* expressing recombinant proteins in the periplasm to non-producing controls [44], [90], [106], [107], [146], [148], [168], [215]. However, none of these studies have compared how *E. coli* responds when producing the same protein in the periplasm and cytoplasm. In this chapter, experiments were performed to address this gap in our knowledge to improve our understanding of cellular responses that are unique to protein secretion and not just due to protein overexpression.

3.2 Aims

- Grow strains expressing a model protein (an anti-TNF scFv antibody fragment) in the cytoplasm and the periplasm (by adding an OmpA signal peptide) under industrially relevant high cell density culture (HCDC) fermentation conditions.
- Analyse samples using transcriptomics and proteomics to look at cellular responses to scFv secretion, secretion system overloading and how cellular responses change over time.

3.3 Results/discussion

3.3.1 Cloning and testing scFv expression

3.3.1.1 Overview of the scFv

An anti-TNF scFv antibody fragment was chosen as a model recombinant protein for this study as antibody fragments are one of the most common therapeutic proteins expressed in the periplasm [1]. Furthermore, this recombinant protein had previously been used in unpublished proteomics work collected by our group. This scFv contains one disulphide bond (two cysteines) and is relatively small (110 amino acids, 12 kDa) (Figure 18). The template for the codon optimized scFv sequence was kindly obtained from Fujifilm Diosynth Biotechnologies FDB.

The scFv was secreted to the periplasm using an N-terminal OmpA signal peptide, as previous studies at FDB identified that this signal peptide was optimal for expression and secretion of the scFv to the periplasm. The OmpA signal peptide translocates proteins via the Sec secretion pathway post-translationally [102], [256]. Adding the OmpA signal peptide added an extra 16 amino acids to the scFv (Figure 18). It is important to consider that addition of this signal peptide could also affect transcription and translation of the scFv, cellular responses and RNA stability, with the 5' terminus playing an important role in transcript degradation [257].

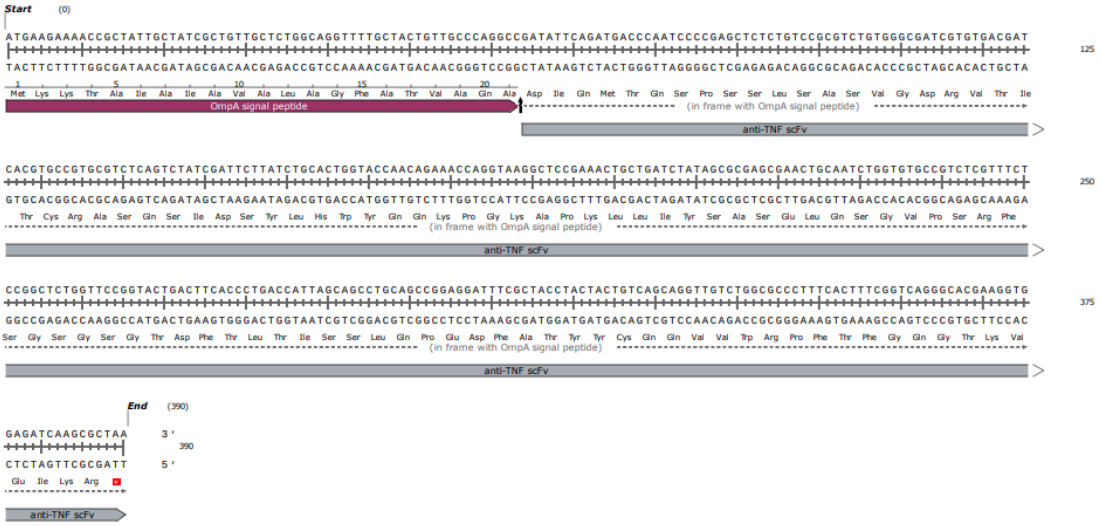


Figure 18. OmpA signal peptide and scFv sequence. Both nucleotide and three letter amino acid sequences are provided. Figure was created using SnapGene.

3.3.1.2 Cloning the scFv with and without a signal peptide

To compare *E. coli* expressing an scFv in the cytoplasm and periplasm, the scFv sequence was cloned into an expression vector with and without the N-terminal OmpA signal peptide. PCR was used to amplify from a template using primers that either included the signal peptide (paveway_11_F and scFv_R) or excluded the signal peptide (scFv_F and scFv_R primers). These PCR products were then cloned into two different expression vectors from FDB: paveway 11 (strong inducible promoter) and paveway 29 (weak inducible promoter) using NdeI and XhoI restriction sites. Successful cloning was then verified using colony PCR as can be seen by the presence of 600 bp and 550 bp bands that represent the scFv with and without the OmpA signal peptide (Figure 19). Plasmids were then further verified using sanger sequencing.

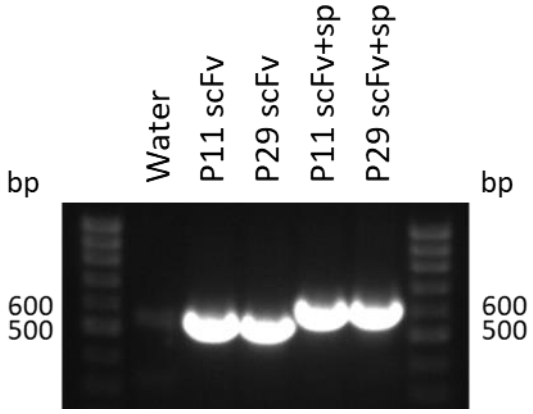


Figure 19. Colony PCR verifying cloning of our model protein (scFv) with and without the OmpA signal peptide (sp) into paveway 11 (P11) and paveway 29 (P29) expression vectors using Paveway_11_F and T7_terminator_R primers. Successful cloning of the scFv leads to amplification of a 603 bp product with the signal peptide and 550 bp product without. Water was used as a negative control however there was a small amount of DNA carryover from adjacent wells.

3.3.1.3 Testing scFv expression – shake flasks

Once successfully cloned, expression and secretion of the scFv were verified by transforming plasmids into an overexpression strain used by our industrial collaborators to ensure relevance to their processes (*E. coli* W3110 Δ*ompT*, FDB) and inducing expression under shake flask conditions (see methods section 2.2.4). Samples were induced with 0.125 mM IPTG, the same concentration as had been used in a proteomics study conducted by our group. Whole cells were analysed using SDS-PAGE, before induction (negative control), 3 hours post-induction and in overnight samples (Figure 20).

The scFv was successfully expressed and secreted when using the paveway 11 expression vector (P11) as can be seen by the presence of the 12 kDa bands (scFv without signal peptide) 3 hours post-induction and overnight but not in the pre-induced negative controls (Figure 20). A 14 kDa band was also present in samples expressing the scFv in the periplasm. This band represents the unprocessed scFv that still contains the OmpA signal peptide, indicating that some of the scFv is accumulating in the cytoplasm. The scFv represents a large portion of the *E. coli* proteome as the intensity of the scFv bands represents around 28% of total proteins when inducing expression of the scFv in the

cytoplasm and 42% of total proteins when expressing the scFv in the periplasm based on image analysis of band intensities (see methods section 2.4.2.4).

No 12 or 14 kDa bands that represent the scFv were detectible when using the paveway 29 plasmid to overexpress the scFv (Figure 20). This may be due to the paveway 29 promoter being weaker or due to mutations in the plasmid. The paveway 11 plasmid was therefore used in all following work to overexpress the scFv.

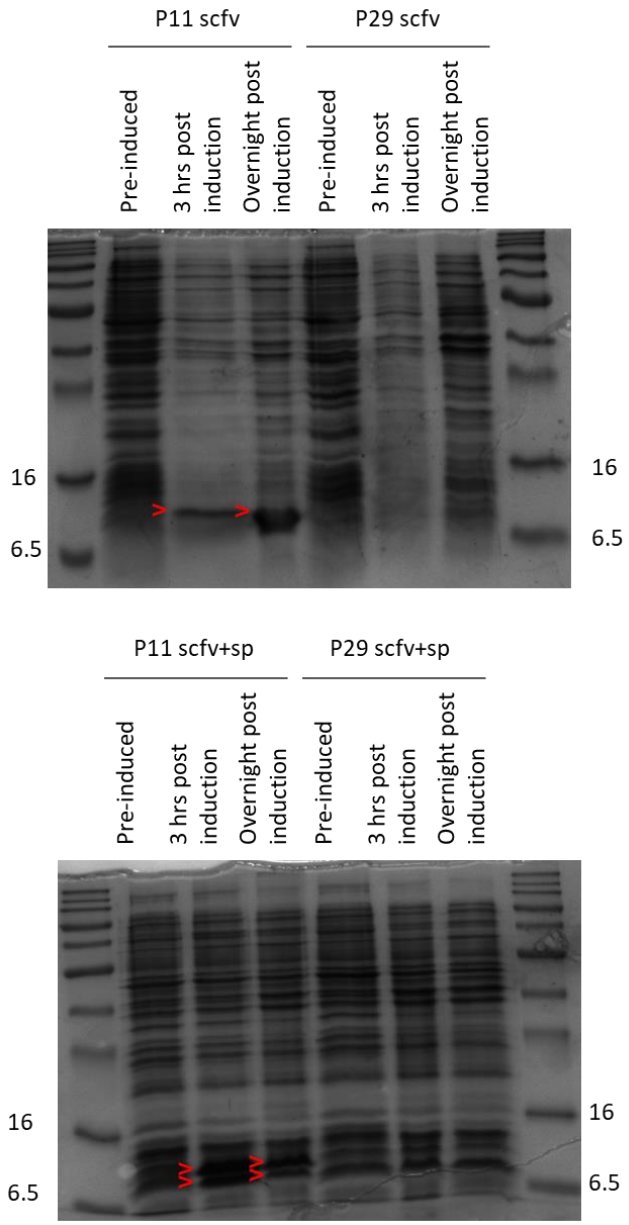


***Figure 20. SDS-PAGE of E. coli whole cells overexpressing the scFv with and without an N-terminal OmpA signal peptide (sp) under shake flask conditions when cloned into the paveway 11 (P11) or paveway 29 (P29) plasmid. Loading was normalized based on bacterial density (OD_{600}). The scFv has a predicted molecular weight of* 14 *kDa* with the signal peptide *and 12 kDa without the signal peptide and is indicated using arrows.* Samples were induced with 0.125 mM IPTG mid exponential phase (OD_{600} 3-4).**

3.3.2 Expression of scFv in the periplasm and cytoplasm under industrially relevant conditions – High Cell Density Cultures - HCDC

3.3.2.1 Introduction

The next step was to express our model protein (scFv) under industrially relevant conditions (HCDC High Cell Density Cultures) at FDB using Ambr250 fed batch fermenters. Samples expressing the scFv in cytoplasm (no signal peptide) and periplasm (OmpA signal peptide) were grown in triplicate. Samples were induced with two different levels of IPTG (0.125 and 0.0125 mM) to determine the effect of protein expression levels and secretion system overloading on cellular responses (Figure 21). Upon induction, samples were taken every 12 hours for a total of 48 hours. One of the replicates (0.0125 mM IPTG periplasmic expression of the scFv) had to be discarded due to a faulty antifoam sensor. Two of the samples expressing the scFv in the cytoplasm (0.0125 mM IPTG) were taken as part of a separate Ambr run. Prior to transcriptomic and proteomic analysis whole cells, supernatant, periplasmic/spheroplast and soluble/insoluble extracts were then extracted and analysed using SDS-PAGE and LabChip capillary electrophoresis to check expression and secretion of the model protein (scFv).

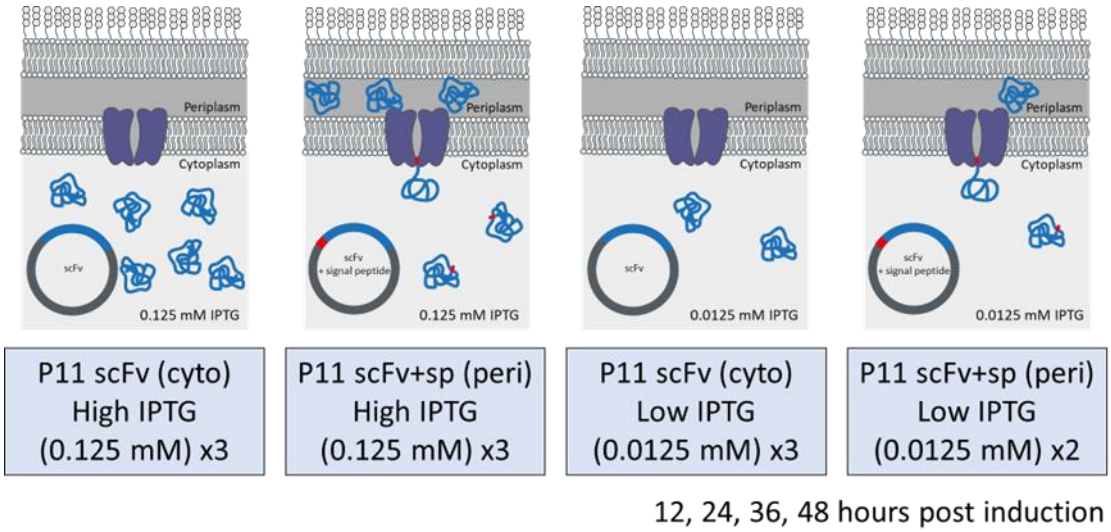


Figure 21. Samples generated from Ambr250 fermentation. The scFv was either expressed in the cytoplasm (cyto – P11 scFv - with no signal peptide) or the periplasm (peri – P11 scFv+sp – with ompA signal peptide). Samples were also either induced with "high" levels of IPTG (0.125 mM) or "low" levels of IPTG (0.0125 mM). Biological triplicates were used for each sample except for P11 scFv+sp 0.0125 mM IPTG which was taken in duplicate due to a faulty bioreactor. Samples were taken 12, 24, 36 and 48 hours post-induction.

3.3.2.2 Monitoring growth

Growth was monitored using absorbance at OD_{600} and dry cell weight. While OD_{600} is a rapid way of monitoring cell growth, it can often be affected by changes in cell physiology and the presence of inclusion bodies which diffract light. Dry cell weight was therefore used for growth curves (Figure 22). OD_{600} growth curves can be found in Appendix 3.

When expressing the scFv in the cytoplasm (P11-scFv), bacteria maintained growth till 36 hours post-induction before entering the stationary phase regardless of the amount of IPTG added (0.125 or 0.0125 mM IPTG) (Figure 22). In contrast, the stationary and death phase occurred earlier in samples overexpressing the scFv in the periplasm (P11-scFv+sp) (24 hours post-induction) when adding a high concentration of inducer (0.125 mM IPTG peri) and between 12-24 hours post-induction, when lower concentrations of IPTG were used (0.0125 mM IPTG peri) (Figure 22). This decrease in growth rate and cell death could be due to multiple factors including cell lysis/leakiness due to overloading of the periplasm and/or the metabolic burden of recombinant protein secretion.

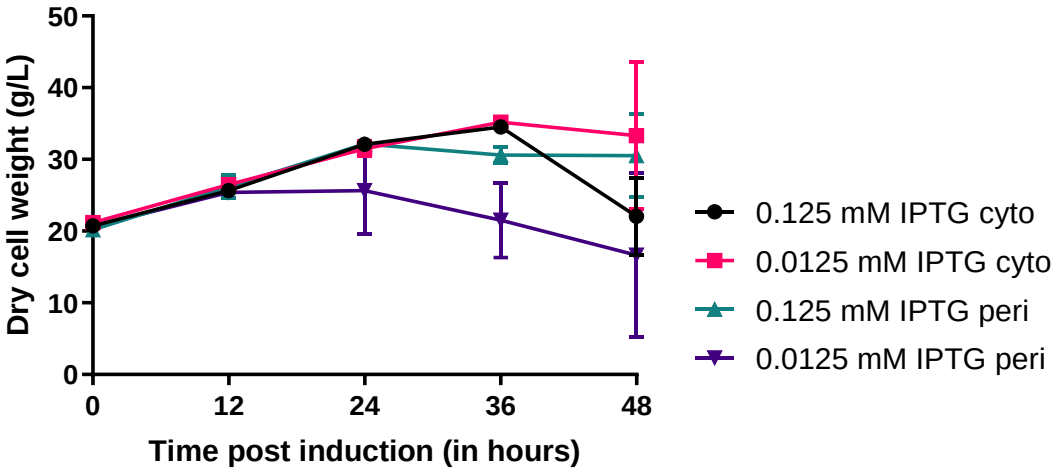


Figure 22. Growth curve of high cell density E. coli cultures expressing the scFv in the cytoplasm (cyto P11-scFv) or periplasm (peri P11-scFv+sp) induced with 0.125 mM IPTG or 0.0125 mM IPTG. Bacterial concentration was monitored using dry cell weight. Error bars = 1 standard deviation, N=3 for all samples except cells inducing expression of the scFv in the periplasm with 0.0125 mM IPTG where N=2. Samples were analysed using a two-way ANOVA. There was a significant decrease in bacterial density when inducing high (0.125 mM IPTG) levels of scFv expression in the periplasm compared to samples expressing the scFv in the cytoplasm (p<0.05).

3.3.2.3 Verifying scFv overexpression – whole cells

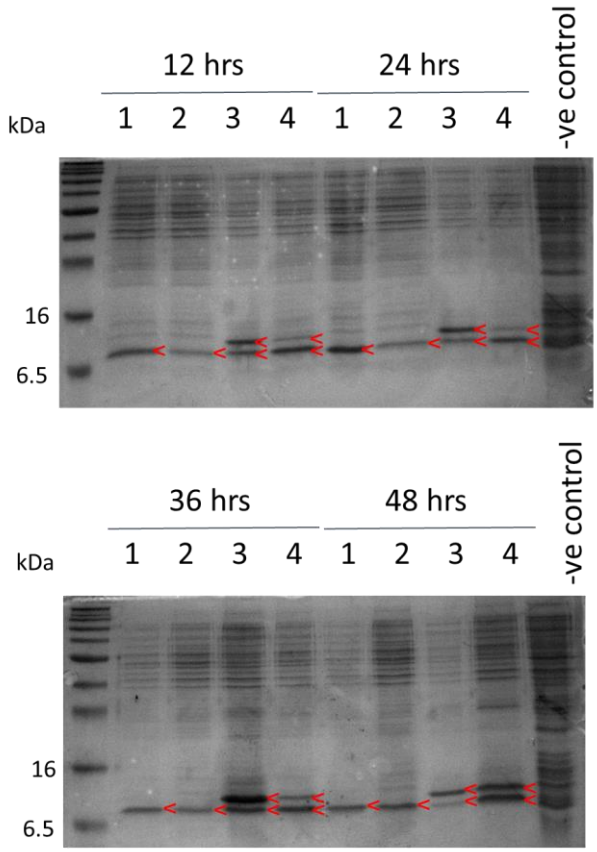
To verify expression of our model protein (scFv), whole cells were analysed using LabChip capillary electrophoresis (Appendix 4) and one replicate from each condition analysed by SDS-PAGE.

The scFv was successfully expressed in the cytoplasm (P11-scFv) as seen by the 12 kDa band (Figure 23, samples 1 and 2) that was not present in the pre-induced negative control. This band was more intense when induced with 0.125 mM IPTG (Figure 23, sample 1) than 0.0125 mM IPTG (Figure 23, Sample 2) within the first 24 hours post-induction, indicating higher yields of scFv.

The secreted scFv was also successfully overexpressed when secreted to the periplasm (P11-scFv+sp) as bands at 12 kDa (processed, no signal peptide) and 14 kDa (unprocessed, contains signal peptide) (Figure 23, sample 3 and 4) were present in these samples that were not found in the pre-induced control. There was a higher proportion of unprocessed scFv (14 kDa) compared to processed scFv (12 kDa) when inducing high levels of scFv expression in the periplasm (Figure 1, sample 3) than when adding lower amounts of inducer

(Figure 1, sample 4). This indicates that the secretion system was probably overloaded and the scFv was accumulating in the cytoplasm in these samples.

Overall, the intensity of scFv bands was more intense when secreting the scFv to the periplasm (Figure 23, sample 3 and 4) than when expressing the scFv in the cytoplasm (Figure 23, samples 1 and 2). This could be due to multiple factors. The additional signal peptide could have affected transcription and translation and stability of the scFv. Furthermore, the scFv may face higher levels of degradation when expressed in the cytoplasm.



1	P11-scFv 0.125 mM IPTG
2	P11-scFv 0.0125 mM IPTG
3	P11-scFv+sp 0.125 mM IPTG
4	P11-scFv+sp 0.0125 mM IPTG

Figure 23. SDS-PAGE of cells overexpressing the scFv in the cytoplasm (P11-scFv) or periplasm (P11-scFv+sp) induced with 0.125 mM IPTG or 0.0125 mM IPTG, 12-, 24-, 36- and 48-hours post-induction. Samples were normalised to the bacterial concentration (OD_{600}). The scFv has a molecular weight of 12 kDa without the signal peptide and 14 kDa with the OmpA signal peptide and is indicated by a red arrow. Pre-induced samples were used as a negative control.

3.3.2.4 Verifying scFv secretion - periplasmic extractions

Osmotic shock was used to extract periplasmic proteins and test for secretion of the scFv to the periplasm. All periplasmic fractions were analysed

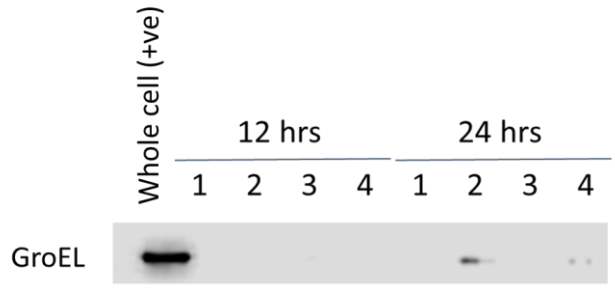
using LabChip capillary electrophoresis (Appendix 5) and one replicate from each condition was analysed using SDS-PAGE (Figure 25). Western blots of GroEL were used to check for contamination of cytoplasmic proteins in the periplasmic fraction as GroEL is only present in the cytoplasm. In the periplasmic fraction, GroEL bands were not detectible or were a lot less intense than the whole cell positive control indicating that there was very little cytoplasmic contamination in the periplasmic fractions (Figure 24). Furthermore, the band profile of the periplasmic fraction was very different to that of cytoplasmic proteins indicating that the periplasmic fraction did not contain large amounts of cytoplasmic proteins (Figure 23, Figure 25).

The scFv was successfully secreted to the periplasm. A band representing the secreted/processed scFv (12 kDa) was present in the periplasmic fraction of samples secreting this protein to the periplasm (Figure 24, samples 3 and 4) and not the negative control. More of the model protein (scFv) was secreted to the periplasm when adding lower amounts of inducer (Figure 24, sample 4) compared to higher levels of inducer (Figure 24, sample 3) as scFv bands were more intense. This is consistent with previous findings that the secretion system was overloaded when inducing high levels of scFv expression (0.125 mM IPTG).

A band representing the 12 kDa scFv was also present periplasmic fraction of samples where the scFv was expressed in the cytoplasm (Figure 24, sample 1 and 2), however, these bands were a lot less intense than samples where the scFv was secreted (Figure 24, samples 3 and 4). The presence of this 12 kDa was probably due to small levels of cell lysis or leakiness when preparing periplasmic extracts.

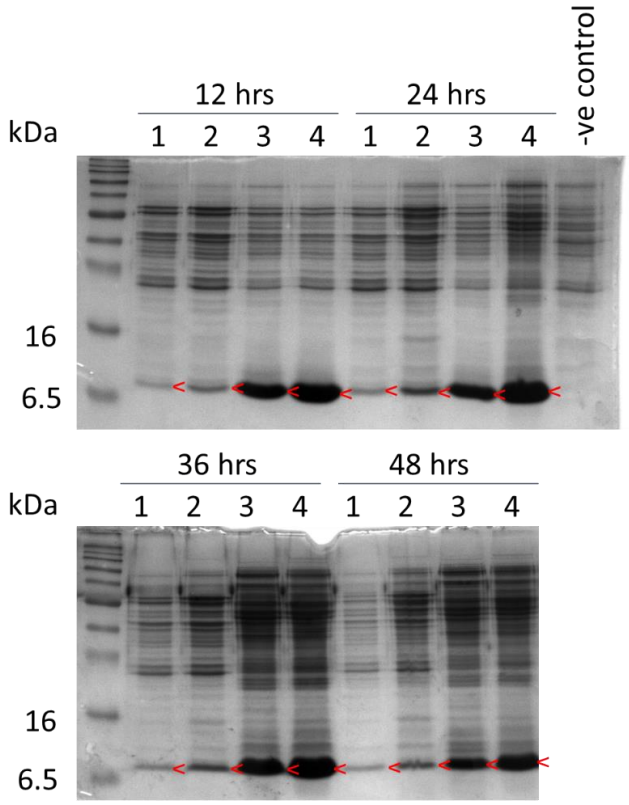
Spheroplasts (cells where the outer membrane and periplasm are removed) were also analysed using SDS-PAGE (Figure 26). The majority of the 12 kDa processed scFv band was no longer present in the spheroplasts of cells secreting the scFv indicating that periplasmic extraction was successful. More unprocessed scFv accumulated in the spheroplast when inducing high amounts of scFv expression (0.125 mM IPTG) (Figure 26, samples 3) compared to lower amounts of scFv expression (0.0125 mM IPTG) (Figure 26, sample 4) which also indicates secretion system overloading. A band at around 12 kDa was also present in the pre-induced negative control which may interfere with the

quantification of scFv bands in these samples. The presence of this band may be either due to leaky expression of the scFv or the presence of a protein with a similar molecular weight in spheroplasts.



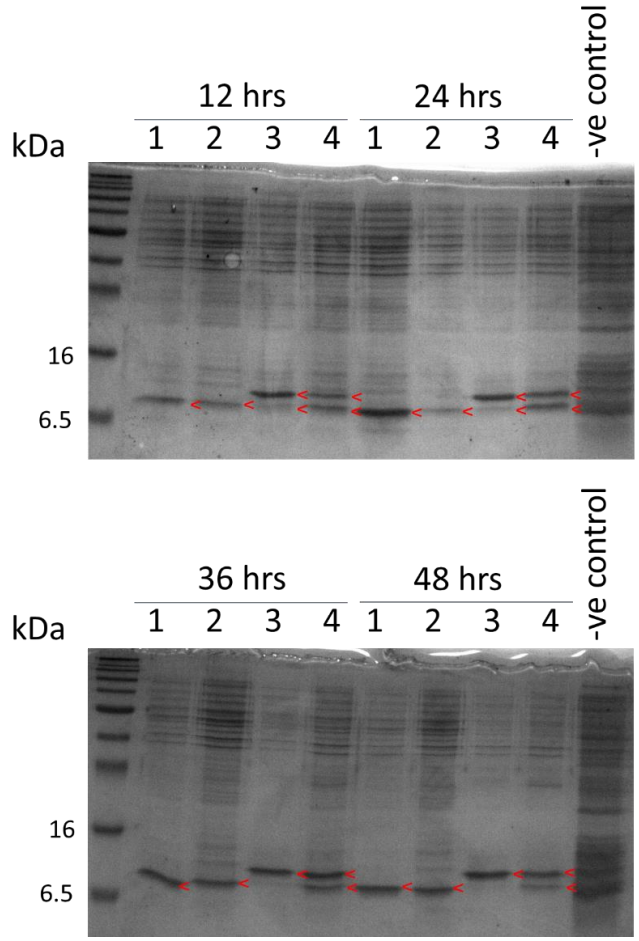
1	P11-scFv 0.125 mM IPTG
2	P11-scFv 0.0125 mM IPTG
3	P11-scFv+sp 0.125 mM IPTG
4	P11-scFv+sp 0.0125 mM IPTG

Figure 24. Western blot of periplasmic extracts detecting the cytoplasmic protein GroEL. The periplasm was extracted from high cell density cultures expressing the scFv in the cytoplasm (P11-scFv) or periplasm (P11-scFv+sp) induced with 0.125 mM IPTG or 0.0125 mM IPTG, 12 hours and 24 hours post-induction. Gel loading was normalised to bacterial density (OD_{600}). Whole cells were used as a positive control. GroEL has a molecular weight of approximately 57 kDa.



1	P11-scFv 0.125 mM IPTG
2	P11-scFv 0.0125 mM IPTG
3	P11-scFv+sp 0.125 mM IPTG
4	P11-scFv+sp 0.0125 mM IPTG

Figure 25. SDS-PAGE of periplasmic extracts from high cell density cultures expressing the model protein scFv in the periplasm (P11-scFv+sp) or cytoplasm (P11-scFv) induced with 0.125 mM IPTG and 0.0125 mM IPTG, 12-, 24-, 36- and 48-hours post-induction. The periplasmic fraction was extracted using cold osmotic shock and was normalised to bacterial concentration OD_{600}. The scFv has a predicted molecular weight of 14 kDa with the signal peptide and 12 kDa without the signal peptide and is indicated by a red arrow. Pre-induced samples were used as a negative control.



1	P11-scFv 0.125 mM IPTG
2	P11-scFv 0.0125 mM IPTG
3	P11-scFv+sp 0.125 mM IPTG
4	P11-scFv+sp 0.0125 mM IPTG

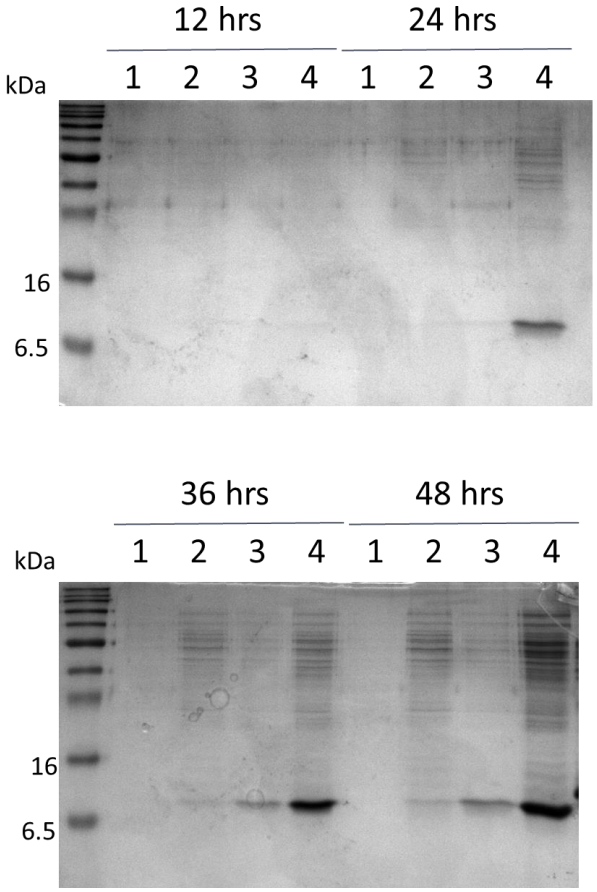
Figure 26. SDS-PAGE of spheroplasts from high cell density cultures expressing the scFv in the cytoplasm (P11-scFv) or periplasm (P11-scFv+sp) induced with 0.125 mM IPTG and 0.0125 mM IPTG, 12-, 24-, 36- and 48-hours post-induction. The concentration of spheroplasts loaded to each lane was normalised to bacterial concentration OD_{600}. The scFv has a predicted molecular weight of 14 kDa with the signal peptide and 12 kDa without the signal peptide and is indicated by a red arrow. Pre-induced samples were used as a negative control.

3.3.2.5 Monitoring cell lysis – supernatant samples

Expression of a large quantity of proteins in the periplasm often leads to cell lysis. Supernatants were therefore analysed to look for the presence of whole cell proteins in the supernatant using LabChip capillary electrophoresis (Appendix 6) and one replicate was analysed using SDS-PAGE (Figure 27).

E. coli lysed 24 hours post-induction when expressing our model protein (scFv) in the periplasm as seen by the presence of whole cell proteins in the supernatant using SDS-PAGE analysis (Figure 27, samples 3 and 4). It is interesting to note that the supernatant has a similar band profile to periplasmic

fractions indicating that cells are leaky rather than lysing (Figure 25, Figure 27). This is consistent with findings from growth curves that saw a decrease in growth 24 hours post-induction (Figure 22). This leakiness is more severe when inducing lower levels of scFv induction (0.0125 mM IPTG) in the periplasm as bands in the periplasmic fraction are more intense (Figure 27, sample 4). This is consistent with previous findings that indicate that the periplasm is more overloaded under these conditions (see section 3.3.2.4). There was however minimal cell lysis when expressing the scFv in the cytoplasm as no protein bands can be seen in the supernatant in these samples other than slight lysis in samples induced with 0.0125 mM IPTG 24 and 48 hours post induction (Figure 27, samples 1 and 2).



1	P11-scFv 0.125 mM IPTG
2	P11-scFv 0.0125 mM IPTG
3	P11-scFv+sp 0.125 mM IPTG
4	P11-scFv+sp 0.0125 mM IPTG

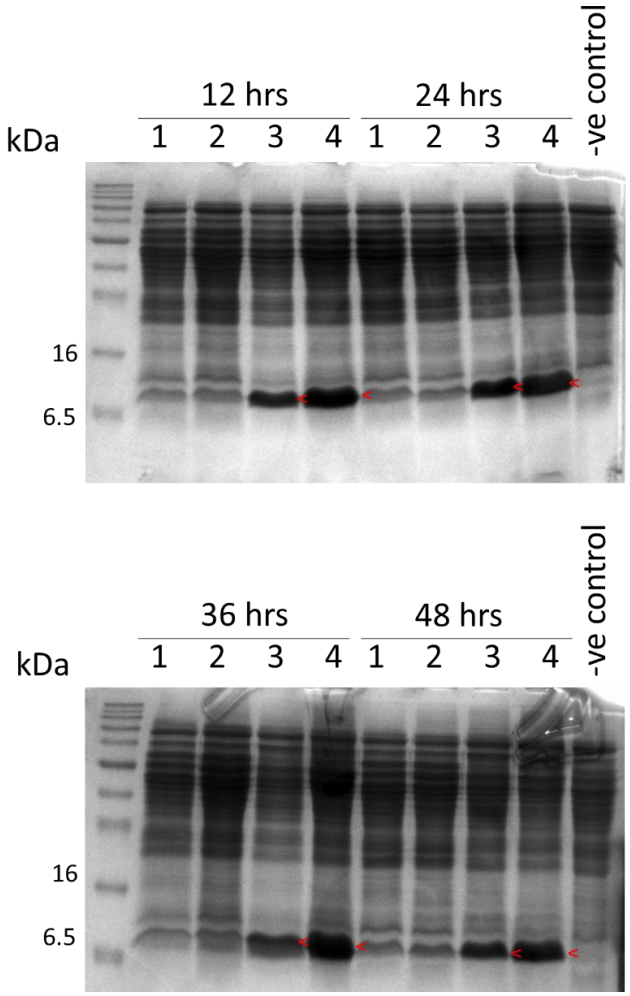
Figure 27. SDS-PAGE of supernatants from high cell density cultures expressing the ScFv in the cytoplasm (P11-scFv) or periplasm (P11-scFv+sp) induced with 0.125 mM IPTG or 0.0125 mM IPTG, 12-, 24-, 36- and 48-hours post-induction. Concentration of supernatant was adjusted based on the bacterial concentration (OD_{600}).

3.3.2.6 Determining correct scFv folding - soluble and insoluble protein extracts

Soluble and insoluble fractions were then extracted to determine if the scFv folded correctly or formed insoluble inclusion bodies. This was done by lysing cells with BugBuster, pelleting out insoluble proteins using high speed centrifugation and analysing samples using SDS-PAGE (Figure 28, Figure 29).

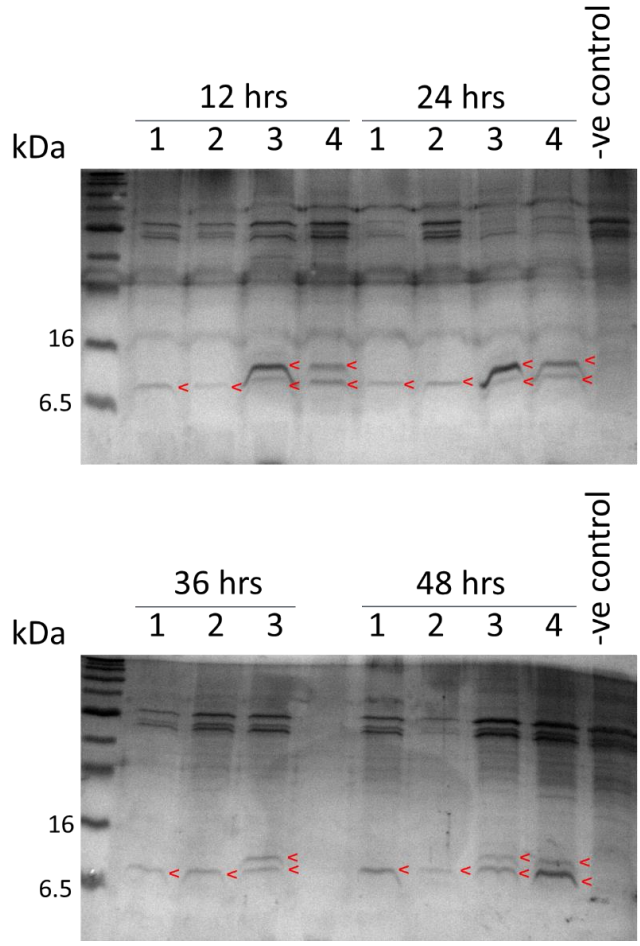
The scFv was soluble when secreted to the periplasm as seen by the presence of 12 kDa band in the soluble extracts which represents the processed scFv (Figure 28, samples 3 and 4). On the other hand, the scFv was insoluble when found in the cytoplasm as seen by the 12 kDa bands in the insoluble fraction, when overexpressing the scFv in the cytoplasm (Figure 29, samples 1 and 2) and the 14 kDa bands when expressing the scFv in the periplasm (Figure 29, samples 3 and 4). This is expected as the scFv contains a disulphide bond which only forms properly in the periplasm (oxidising environment). A 12 kDa band which represents the processed periplasmic scFv, was also present in the insoluble fraction when expressing the scFv in the periplasm, particularly in samples 36- and 48-hours post-induction (Figure 29 samples 3 and 4). This indicates that the scFv may be forming periplasmic inclusion bodies.

One of the few ways of determining if a recombinant protein is forming inclusion bodies in the periplasm is to visualize inclusion bodies using TEM. This was however not possible with samples from Ambr fermentations as these had been frozen. TEM was however conducted on samples overexpressing the scFv in the cytoplasm and periplasm when grown under shake flask conditions. This will be covered in the following section however it is important to consider that these responses may be different to those seen in Ambr fermenters.



1	P11-scFv 0.125 mM IPTG
2	P11-scFv 0.0125 mM IPTG
3	P11-scFv+sp 0.125 mM IPTG
4	P11-scFv+sp 0.0125 mM IPTG

Figure 28. SDS-PAGE of soluble proteins extracted from high cell density cultures expressing an scFv in the cytoplasm (P11-scfv) or periplasm (P11-scfv+sp) induced with 0.125 mM IPTG and 0.0125 mM IPTG, 12-, 24-, 36- and 42-hours post-induction. Cells were lysed using BugBuster, unlysed cells were removed by centrifugation and insoluble proteins were pelleted out using high speed centrifugation. Soluble protein extracts were normalised based on the weight of the bacterial pellet. A pre-induced sample was used as a negative control. The scFv has a predicted molecular weight of 12 kDa without the signal peptide and 14 kDa with the signal peptide and is indicated by a red arrow.



1	P11-scFv 0.125 mM IPTG
2	P11-scFv 0.0125 mM IPTG
3	P11-scFv+sp 0.125 mM IPTG
4	P11-scFv+sp 0.0125 mM IPTG

Figure 29. SDS-PAGE of insoluble proteins extracted from high cell density cultures expressing an scFv in the cytoplasm (P11-scfv) or periplasm (P11-scfv+sp) with 0.125 mM IPTG and 0.0125 mM IPTG at various time points. Cells were lysed using BugBuster, unlysed cells were removed by centrifugation and insoluble proteins were pelleted using high speed centrifugation. Insoluble protein extracts were normalised based on bacterial pellet weight. A pre-induced sample was used as a negative control. The scFv has a predicted molecular weight of 12 kDa without the signal peptide and 14 kDa with the signal peptide and is indicated by a red arrow.

3.3.2.6.1 Detecting periplasmic inclusion bodies in shake flasks using TEM

Cross sectional Transmission Electron Microscopy (TEM) was also used to look for the presence and location of inclusion bodies in the cell when secreting the scFv to the periplasm under shake flask conditions. Inclusion bodies are electron dense and therefore form dark patches in the cytoplasm or periplasm when using TEM [93], [95], [258].

Three different types of samples were analysed using TEM, an empty vector control (Paveway 11), cells expressing the scFv in the cytoplasm (P11-

scFv) and cells expressing the scFv in the periplasm as an OmpA signal peptide had been added (P11 scFv+sp). Bacteria were grown under shake flask conditions and samples were taken 3 hours post-induction with 0.125 mM IPTG and were immediately processed for cross sectional TEM. Biological triplicates were taken for each sample. SDS-PAGEs of whole cell and supernatant samples can be found in (Appendix 7) which verified that the scFv was expressed and secreted. Images were taken at a high magnification (6800x) to look at specific cellular structures and a lower magnification (2900x) to present a less biased view of cellular morphology (Figure 30).

Expression of the scFv in the cytoplasm (P11 scFv) led to the accumulation of cytoplasmic inclusion bodies, as can be seen by the presence of dark spots of aggregated protein that are not found in the uninduced control (Figure 30). This is consistent with SDS-PAGE analysis of insoluble proteins (Figure 29) and is expected as the scFv contains a disulphide bond and therefore does not fold properly in the cytoplasm. Cytoplasmic inclusion bodies primarily form at the poles of the cell [93], [258] which was also observed when overexpressing the scFv in the cytoplasm (Figure 30). Inclusion bodies were a lot smaller than those often reported in the literature [93], [258], however, this is probably due to low levels of recombinant protein overexpression.

Expression of the scFv in the periplasm (P11 scFv+sp) also led to an accumulation of inclusion bodies in the cytoplasm as seen by the dark electron dense spots that are not found in the empty vector control (Figure 30). These inclusion bodies form at the poles of cells which is consistent with previous studies [95]. This is also consistent with previous findings in section 3.3.1.3 that indicate that some of the scFv remains in the cytoplasm when the periplasmic scFv is overexpressed. The location of these inclusion bodies is however different when secreting the scFv to the periplasm. The inclusion bodies primarily form around the membrane rather than the cytoplasm when secreted. This could be a consequence of cell leakiness/lysis but could also be due to overloading of the secretion system. Indeed, inclusion bodies have been shown to form at the membrane when expressing an export-incompetent misfolded protein via the TAT pathway when analysing cells [44] and when overloading the secretion system by expressing a heavy chain mAb [259]. Other studies

expressing a different periplasmic recombinant protein did not observe inclusion bodies near the membrane and instead inclusion bodies were found at the poles of cells [95]. This however may be due to higher levels of recombinant protein overexpression and the fact that the overexpressed protein was more prone to form inclusion bodies.

TEM images indicate that insoluble inclusion bodies are forming in the periplasm when secreting the scFv as seen by the presence of electron dense areas in the periplasm in several cells (Figure 31). The scFv may not be very stable and/or scFv expression may be overwhelming proteins involved in periplasmic protein folding. Indeed, multiple studies have identified that proteins can form inclusion bodies in the periplasm [66], [93]–[95]. This is consistent with previous analysis of insoluble fractions that indicated that some of the processed scFv remains insoluble when in the periplasm (see section 3.3.2.6) however this primarily occurred later on in the fermentation runs (48 hrs post induction) (Figure 29 sample 3) which is different to samples analysed using TEM that were taken 3 hours post induction as shake flask cultures do not maintain growth for as long as fed batch fermenters. Periplasmic inclusion bodies may therefore be more prominent later in growth.

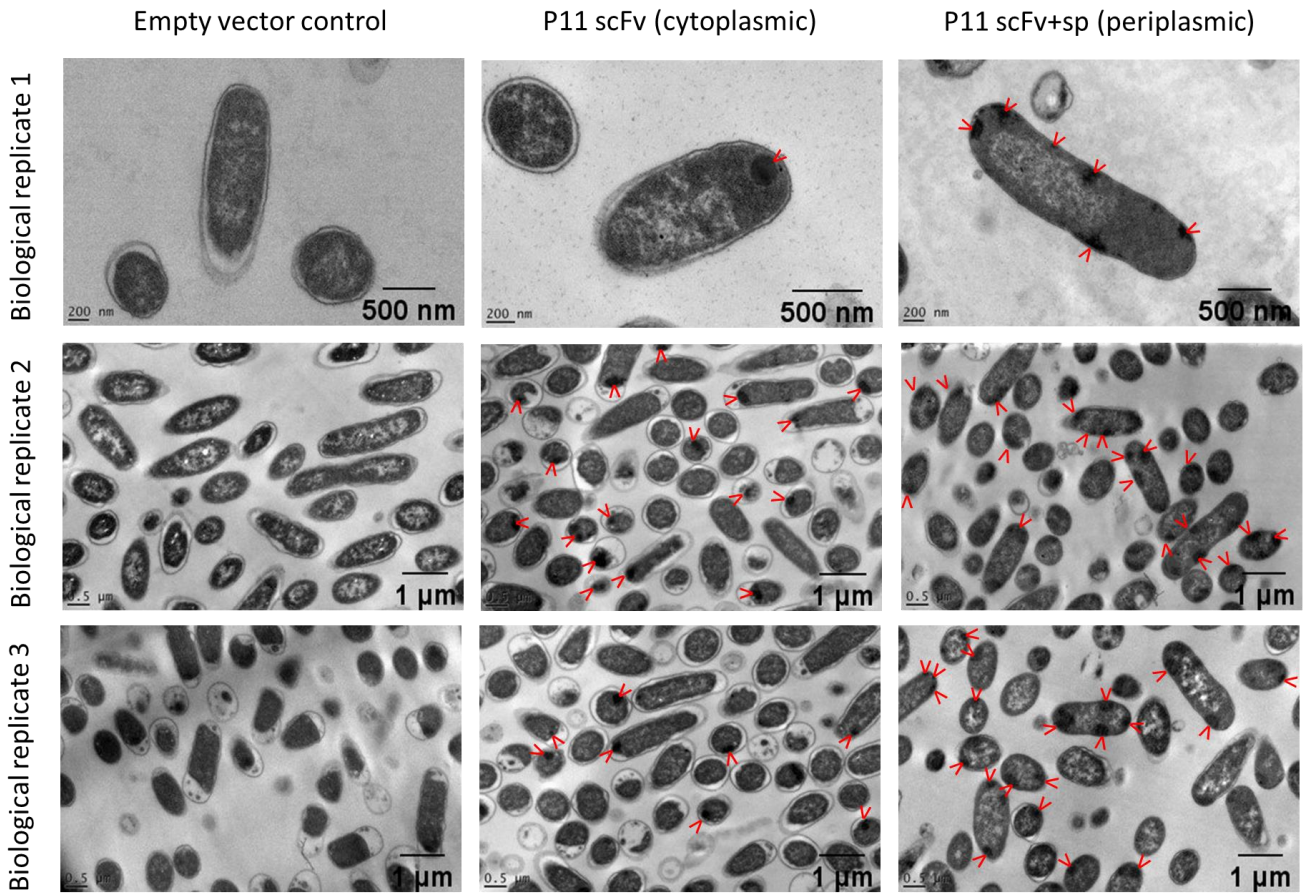


Figure 30. Cross sectional Transmission Electron Microscopy (TEM) E. coli expressing an scFv in the cytoplasm (P11 scFv), the periplasm (P11 scFv+sp) and a non-producing control (empty vector-Paveway 11). Samples were taken 3 hours post-induction with 0.125 mM IPTG under shake flask conditions. Biological replicates were taken for each sample. Inclusion bodies are electron dense and are indicated by a red arrow.

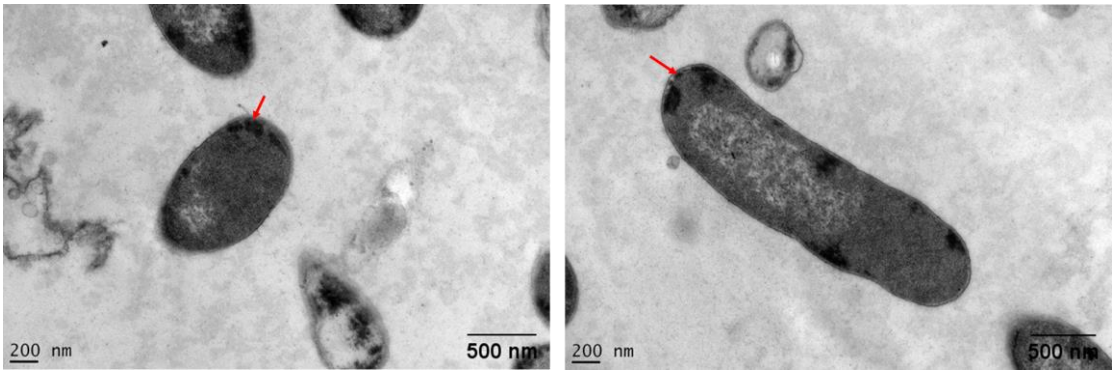


Figure 31. Cross sectional Transmission Electron Microscopy (TEM) showing potential periplasmic inclusion bodies when overexpressing the scFv in the periplasm (P11 scFv+sp). Samples were taken 3 hours post-induction with 0.125 mM IPTG and were grown in shake flask conditions. Periplasmic inclusion bodies are indicated by the red arrow.

3.3.2.7 Expression of scFv in the periplasm and cytoplasm under industrially relevant conditions – Conclusions

The model protein (scFv) was successfully expressed in the cytoplasm (without signal peptide) and periplasm (with signal peptide) under industrially relevant fermentation conditions. As expected, the scFv was insoluble in the cytoplasm and soluble in the periplasm as it contains a disulphide bond. However, TEM and analysis of insoluble protein fractions, indicate that some of the scFv may be forming insoluble inclusion bodies in the periplasm. Inducing scFv expression with a lower concentration of IPTG (0.0125 mM IPTG) led to increased yields of the scFv in the periplasm compared to higher levels of inducer (0.125 mM) that probably overloaded the secretion system and led to a higher level of scFv accumulation in the cytoplasm. Growth remained constant in cells expressing the scFv in the cytoplasm however in cells overexpressing the scFv in the periplasm cells growth decreased 24 hours post-induction, particularly in samples induced with lower concentrations of IPTG (0.0125 mM). This correlates with the increase in cell leakiness and lysis in these samples, probably due to overloading of the periplasm. These samples were further analysed using transcriptomics and proteomics to determine how cellular responses differ between samples, which will be discussed in the following section.

3.3.3 OMICs analysis of *E. coli* overexpressing an scFv in the cytoplasm and periplasm

3.3.3.1 Introduction

Samples expressing our model protein (scFv) in the cytoplasm and periplasm with different levels of inducer (0.125 and 0.0125 mM IPTG) under industrially relevant conditions were further analysed using transcriptomics (RNA-seq) and proteomics (iTRAQ) to look at the effect of scFv secretion, secretion system overloading and how cellular responses change over time. All samples were analysed 24 hours post-induction as bacteria had not yet entered the death phase (Figure 32). Periplasmic samples were also analysed 12 hours post-induction using transcriptomics to determine how cellular responses changed over time as stress from protein production and secretion accumulated (Figure 32).

The following section will start by giving an overview of the quality of transcriptomic and proteomic data. Then the number and overlap between differentially expressed genes and proteins between different conditions will be covered. Finally, the function of differentially expressed genes will be discussed by grouping differentially expressed genes based on their functional activity and cellular location.

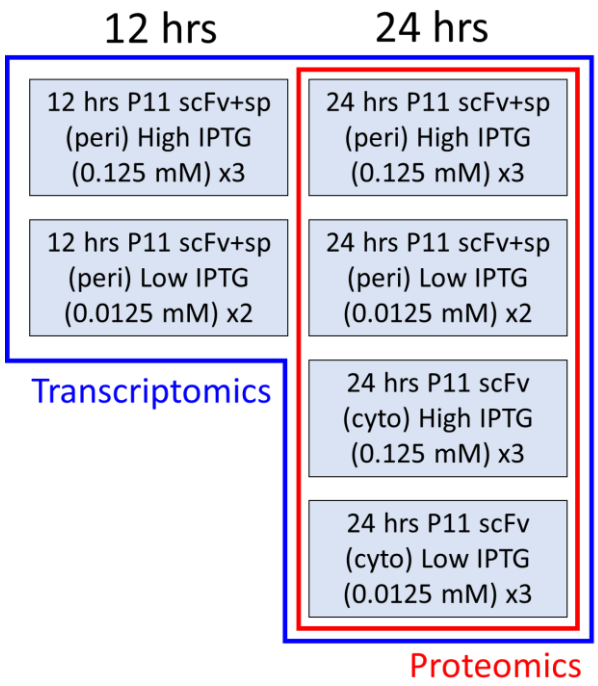


Figure 32. Samples analysed using transcriptomics (circled in blue) or proteomics (circled in red). The scFv was either expressed in the periplasm (peri scFv+sp) or cytoplasm (cyto 0.125 mM), induced with high (0.125 mM) or low (0.0125 mM) concentrations of IPTG and taken 12- or 24-hours post-induction. Biological replicates (x3) or duplicates (x2) were taken for each sample.

3.3.3.2 Quality of Transcriptomics (RNA-seq) data

Transcriptomics was conducted on *E. coli* expressing our model protein (scFv) in the cytoplasm or periplasm induced with both 0.125 mM IPTG and 0.0125 mM IPTG, 12- and 24-hours post-induction (Figure 32). All samples were grown under industrially relevant fermenter conditions in triplicate except those expressing the scFv in the periplasm with low induction which was conducted in duplicate (Figure 32). RNA was successfully extracted from all samples, with concentrations ranging between 179 ng/µL and 984 ng/µL and the OD260/280 remaining between 1.8-2.2. Further quality analysis (Qubit and capillary gel electrophoresis), library preparation, Illumina sequencing and differential expression analysis were conducted by Genewiz.

Illumina sequence quality was verified using FastQC [240] and visualised using mutiQC [241]. Sequence quality was high in all samples however overrepresented sequences were present in all samples (Appendix 8, Appendix

9). This is probably due to the high abundance of reads containing the overexpressed scFv

Reads were then aligned to the *E. coli* W3110 genome as well as the scFv sequence with and without the signal peptide using STAR aligner v.2.5.2b [244] and FeatureCounts [245]. On average, reads were assigned to 4153 of the 4227 known coding regions in *E. coli* W3110 (around 98.3% of coding regions). This indicates that sufficient sequencing depth was used. However, small non coding RNA (sRNA) were not identified as RNA extraction kits are not able to effectively isolate RNA that is smaller than 200 nt [260].

Hierarchical clustering (Figure 33) and PCA analysis (Figure 34) were used to determine whether biological replicates clustered. Hierarchical clustering indicated that all biological replicates expressing the scFv in the periplasm clustered well when induced with different concentrations of IPTG and at different time points with no major outliers (Figure 33). Samples expressing the scFv in the cytoplasm also clustered however there was less of a distinction between samples induced with high and low concentrations of IPTG (0.125 mM and 0.0125 mM IPTG) (Figure 33). All replicates however clustered well when analysed through PCA analysis as replicates showed very little variance and no major outliers were identified (Figure 34).

Differentially expressed genes were then identified using DESeq2. A full list of differentially expressed transcripts can be found in the supplementary Excel spreadsheet "DGE_Transcriptomics" Surprisingly no differences in expression of the scFv were seen between samples even when inducing different levels of scFv expression (0.125 or 0.0125 mM IPTG). This could be due to issues with dynamic range as on average 19% of all mapped counts contained sequences of the scFv. Further analysis of differentially expressed genes will be covered in the following section.

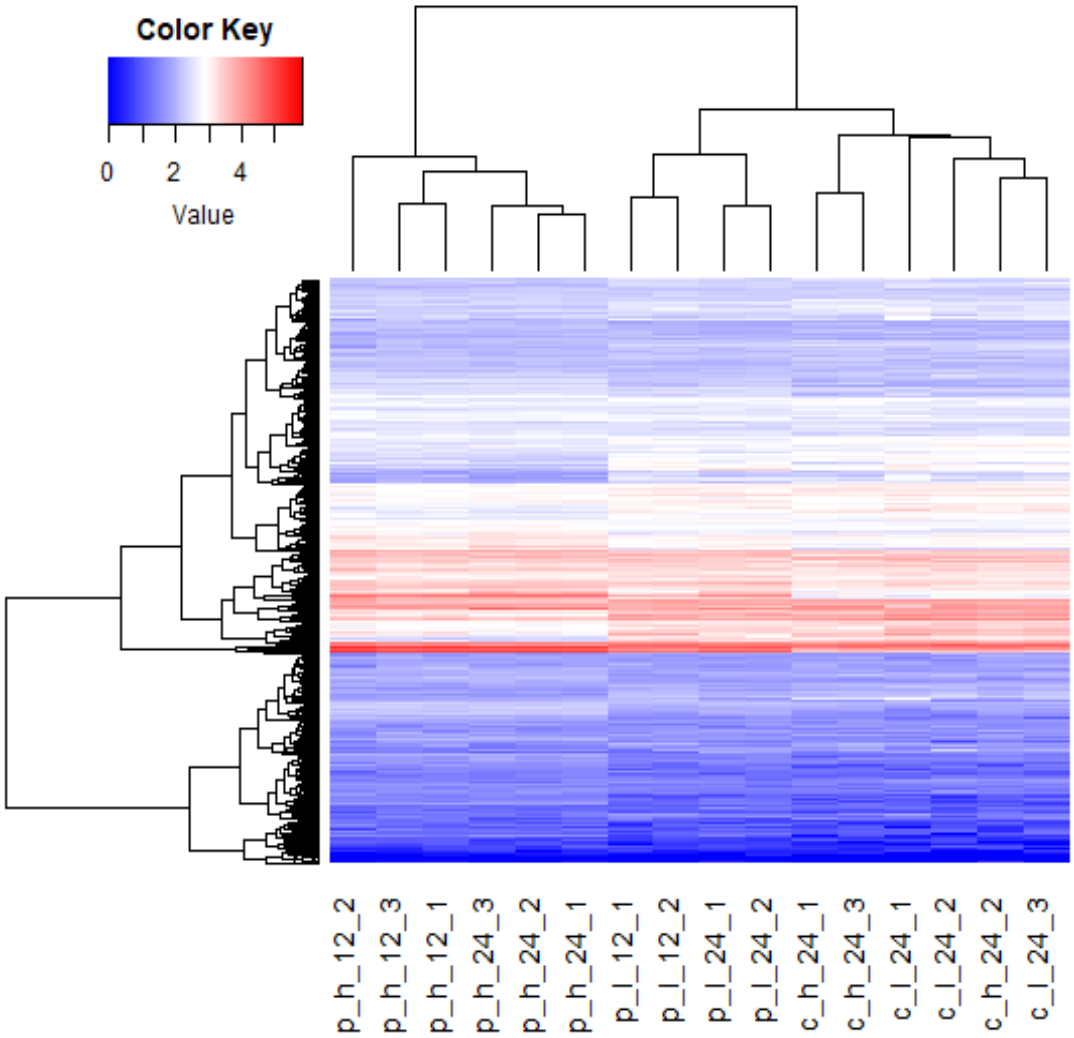


Figure 33. Heatmap of transcriptomics data of E. coli overexpressing an scFv in the periplasm (p) and cytoplasm (c) induced with high (h) 0.125 mM and low (l) 0.0125 mM IPTG concentrations, 12 hrs (12) and 24 hrs (24) post-induction. Heatmap was generated based using the heatmap.2() function in the gplots R package [247]***. Normalized counts were log10 transformed. ScFv counts were not included as these are in such high abundance. Samples were separated using hierarchal clustering using the UPGMA method.***

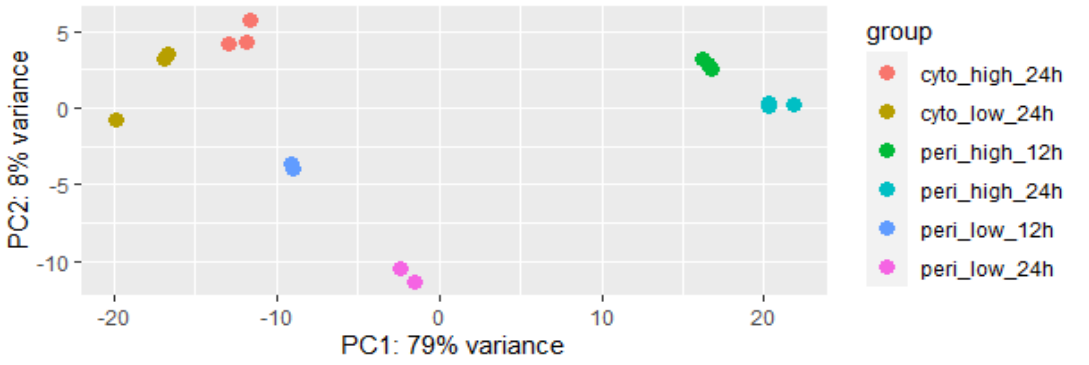


Figure 34. PCA analysis of transcriptomics data. Samples were E. coli overexpressing an scFv in the periplasm (peri) and cytoplasm (cyto) induced with high (0.125 mM IPTG) and low (0.0125 mM IPTG) levels of IPTG, 12 hrs and 24 hrs post-induction. Biological replicates are represented by dots of the same colour. PCA analysis was conducted using the plotPCA() function of the DESeq2 R package [246]***. Counts were log2 transformed.***

3.3.3.3 Quality of Proteomics (iTRAQ) data

Proteomics (iTRAQ) was conducted on samples expressing the scFv in the cytoplasm (no signal peptide) and periplasm (OmpA signal peptide) that were induced with different concentrations of IPTG (0.125 mM IPTG and 0.0125 mM IPTG). Unlike transcriptomics samples that were taken at two timepoints, proteomics samples were only taken 24 hours post-induction (Figure 32). Although technical replicates existed for some of these samples, only technical duplicates were analysed due to the limited number of iTRAQ tags (8-plex). Technical replicates could have been pooled, however, this reduces any information about potential variation between samples [261].

Proteins were extracted from each sample, quantified using a Bradford assay and analysed SDS-PAGE gel to check for equal loading. Samples were then reduced, alkylated, digested with trypsin and analysed using SDS-PAGE to check for successful digestion (Appendix 10) before being tagged with 8-plex iTRAQ tags following the strategy represented in Figure 35. Tagged peptides were pooled, fractionated using HPLC and analysed using tandem mass spectrometry.

Figure 35. Figure representing iTRAQ tagging strategy. iTRAQ tags are indicated by their mass to charge ratio (113, 114, 115, 116, 117, 118, 119, 121). Samples expressing the scFv in the cytoplasm are indicated by scFv, and those secreting the scFv to the periplasm (with an OmpA signal peptide) by scFv+sp. The quantity of inducer was either high (0.125 mM IPTG) or low (0.0125 mM IPTG). Numbers 1-3 indicate specific biological replicates.

Proteins were then quantified using MaxQuaunt. Peptides from 1983 proteins were detected which is equivalent to around 47% of the *E. coli* W3110 proteome. This is expected as mass spectrometry is unlikely to identify the entire proteome with previous studies identifying around 35% of the *E. coli* proteome. Coverage of the *E. coli* proteome can be improved by improving solubilisation of proteins, further fractionating samples and using alternative proteases to trypsin [262].

The scFv was identified in all proteomics samples. There was a significant increase in expression of the scFv when it was secreted to the periplasm compared to samples expressing the scFv in the cytoplasm when induced with both 0.125 and 0.0125 mM IPTG based on differential gene expression using ProteoSign. This is consistent with the SDS-PAGE gel of these samples (Figure 23) and could also be explained by the fact that the scFv was primarily soluble when secreted to the periplasm and insoluble when produced in the cytoplasm (Figure 28 Figure 29) and may not have been sufficiently solubilised during proteomic sample preparation. Surprisingly there was no significant difference in scFv levels when secreting the recombinant protein to the periplasm. There was also no significant difference in scFv levels when adding different concentrations of IPTG 0.125 mM or 0.0125 mM IPTG. This is consistent with transcriptomics data but inconsistent with SDS-PAGE gels analysis and could be due to issues with dynamic range

PCA analysis indicates that all biological replicates except samples expressing low levels of the scFv in the cytoplasm cluster well (Figure 36). It is difficult however to determine any outlier as replicates were only conducted in duplicate. This could also be because samples were generated as part of different fermentation runs. Samples also clustered well based on hierarchical clustering

(Figure 37). Samples where the scFv was expressed at high levels in the periplasm particularly clustered separately from all other samples indicating that their expression profiles were particularly distinct from all samples.

Samples were then analysed using differential gene expression analysis using ProteoSign. A list of differentially expressed proteins can be found in the supplementary Excel spreadsheet "DGE proteomics". No fold change threshold was used to differentially expressed genes as iTRAQ compresses fold change data [184], [187]–[189]. ITRAQ does, however, provide accurate information about fold change direction and therefore, proteomic fold changes will not be discussed in this study and instead, proteins will only be described as being upregulated or downregulated.

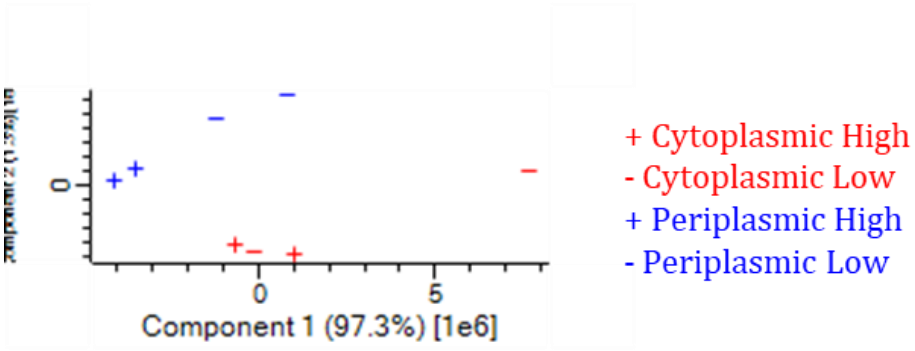


Figure 36. PCA analysis of proteomic data. Dots of the same shape and colour represent biological replicates of E. coli overexpressing an scFv in the periplasm and cytoplasm and induced with high (0.125 mM IPTG) and low (0.0125 mM IPTG) levels of IPTG 24 hrs post-induction. PCA analysis was conducted using Perseus.

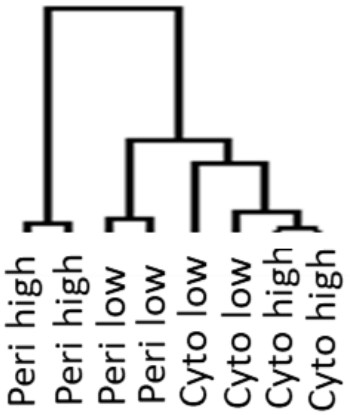


Figure 37. Hierarchical clustering of proteomics data from E. coli samples overexpressing an scFv in the periplasm (peri) and cytoplasm (cyto) induced with high (0.125 mM) and low (0.0125 mM IPTG) level of IPTG 24 hrs post-induction. Hierarchical clustering was conducted using Perseus.

3.3.3.4 Overview – Differentially expressed genes

Multiple comparisons were made as part of this study which can be grouped into 3 themes: cellular responses to scFv secretion, cellular responses to secretion system overloading and changes in cellular responses over time. In the following section, we will provide an overview of differentially expressed genes, discuss the overlap between cellular responses between different conditions tested and the overlap between transcriptomic and proteomic data.

3.3.3.4.1 Cellular responses to scFv secretion to the periplasm

3.3.3.4.1.1 Introduction

To determine the effect of secreting our model protein (scFv) to the periplasm, cells expressing the scFv in the cytoplasm and the periplasm were compared using proteomics and transcriptomics. When expressed in the cytoplasm, the scFv formed insoluble inclusion bodies whereas when the scFv was secreted to the periplasm the scFv was primarily soluble, with some of the scFv remaining in the cytoplasm due to secretion system overloading. This comparison was conducted under two different conditions: induction with a high concentration of IPTG (0.125 mM) and induction with a lower concentration of IPTG (0.0125 mM IPTG). Higher levels of induction led to lower levels of expression of the scFv in the periplasm and higher levels of scFv accumulating in the cytoplasm indicating secretion system overloading.

3.3.3.4.1.2 Transcriptomics overview

Secretion of the scFv to the periplasm also had a significant effect on the *E. coli* transcriptome based on differential gene expression analysis. Secretion of the scFv led to upregulation of 832 and 315 genes and downregulation of 812 and 276 genes when induced with 0.125 mM IPTG and 0.0125 mM IPTG respectively (Figure 38). This indicates that induction of high levels of scFv secretion and secretion system overloading had a more significant impact on the transcriptome. There was however an overlap between differentially expressed transcripts with 241 genes being upregulated and 194 downregulated upon scFv secretion regardless of the concentration of IPTG added (0.125 or 0.0125 mM) and the level of secretion system overloading (Figure 39). A full list of differentially expressed transcripts can be found in the supplementary Excel spreadsheet "DGE_Transcriptomics"

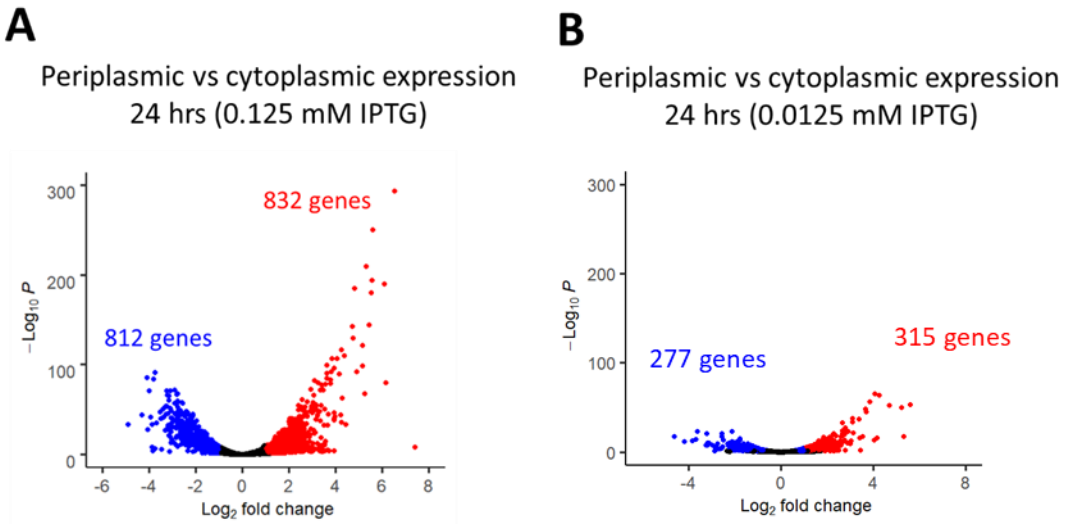


Figure 38. Volcano plots of transcriptomic data comparing the expression of cells secreting a scFv to the periplasm or expressing a scFv in the cytoplasm induced with 0.125 mM IPTG (A) or 0.0125 mM IPTG (B). Black dots represent genes that are not differentially expressed (p value > 0.05, log2 fold change between -1 and 1). Red dots represent genes that are upregulated and blue downregulated when secreting the scFv to the periplasm. Plots were generated using the EnhancedVolcano R package [248].

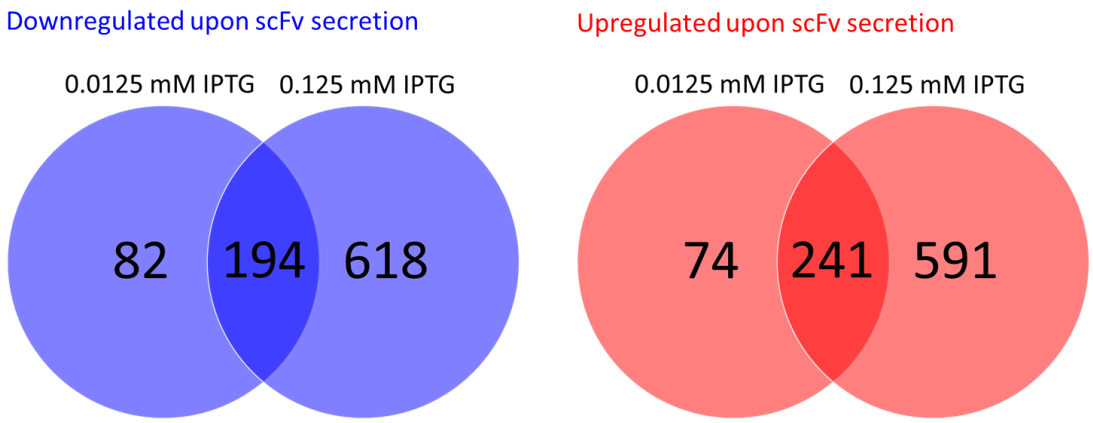


Figure 39. Venn diagram representing the overlap between transcripts upregulated and downregulated upon scFv secretion when induced with different concentrations of IPTG (high (0.125 mM IPTG) and low (0.0125 mM IPTG).

3.3.3.4.1.3 Proteomics overview

Secretion of the scFv to the periplasm also had a significant effect on the *E. coli* proteome. Secretion of the scFv led to an upregulation of 50 and 21 proteins and a downregulation of 32 and 9 proteins when induced with 0.125 mM IPTG and 0.0125 mM IPTG respectively (Figure 38 – fold change data is compressed and therefore only indicates if a protein is upregulated or downregulated). Inducing scFv secretion with high concentrations of IPTG (0.125 mM) has more of an effect on the cell than lower amounts of induction, which is consistent with transcriptomics data. There was however an overlap between differentially expressed proteins with 10 upregulated proteins (YpeB, YciW, BssS, PrpC, PotG, Tig, ClpA, BioD, FhuA, HemF) and 1 downregulated protein (YdcA) when secreting the scFv to the periplasm regardless of the amount of IPTG added and level of secretion system overloading (0.125 mM or 0.0125 mM IPTG). A full list of differentially expressed proteins can be found in the supplementary Excel spreadsheet "DGE_Proteomics".

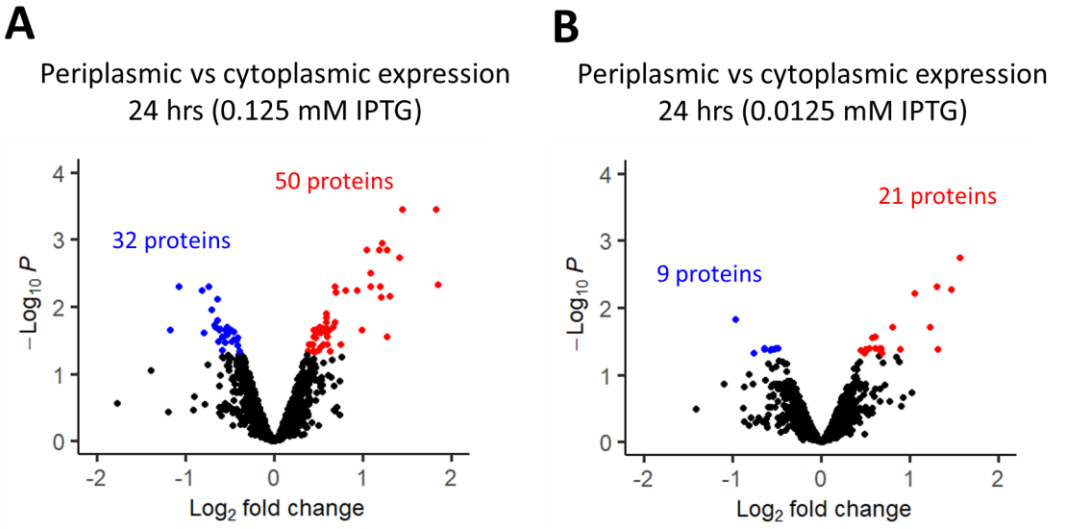


Figure 40. Volcano plots of proteomic data comparing cells secreting an scFv to the periplasm or expressing an scFv in the cytoplasm induced with 0.125 mM IPTG (A) or 0.0125 mM IPTG (B). Black dots represent genes that are not differentially expressed (adjusted p value > 0.05, log2 fold change). Red dots represent proteins that are upregulated and blue downregulated when secreting the scFv to the periplasm. iTRAQ is semi-quantitative and provides accurate information about whether proteins are upregulated or downregulated but not an accurate representation of fold change as this data is compressed. Therefore, differentially expressed genes were not identified using a fold change cut off. Plots were generated using the EnhancedVolcano R package [248].

3.3.3.4.1.4 Overlap between transcriptomics and proteomics

There was an overlap between transcriptomics and proteomics data with 25 genes being upregulated and 9 downregulated in common when inducing high levels of scFv secretion (0.125 mM IPTG) (Figure 37) and 6 upregulated and 3 downregulated in common when secreting low levels of scFv (0.0125 mM IPTG) (Table 37). This is expected as transcript and protein levels often do not correlate due to differences in protein stability. Furthermore, transcriptomics data involved analysing triplicate samples whereas proteomics data was only analysed duplicates and therefore had less statistical power. It is also important to consider that 98% of the transcriptome was identified using RNA-seq whereas only 47% of the proteome was identified using mass spectrometry [263]. Information about whether proteins encoded by differentially expressed transcripts were identified using mass spectrometry can be found in the supplementary Excel spreadsheet "DGE_Transcriptomics". The function of

differentially expressed genes and proteins will be covered in the following sections.

Gene name	Uniprot ID	Log2 fold change transcriptomics	Significant changes proteomics	Function
ibpB	P0C058	6.51713	upregulated	heat shock chaperone
prlC	P27298	5.57725	upregulated	oligopeptidase A
rpoD	P00579	5.55379	upregulated	RNA polymerase, sigma 70 (sigma D) factor
ibpA	P0C054	5.53507	upregulated	heat shock chaperone
uxaB	P0A6L7	5.15683	upregulated	altronate oxidoreductase, NAD-dependent
yibA	P0ADK6	3.93945	upregulated	lyase containing HEAT-repeat
clpX	P0A6H1	3.46018	upregulated	ATPase and specificity subunit of ClpX-ClpP ATP-dependent serine protease
uxuA	P24215	3.43555	upregulated	mannonate hydrolase
mutM	P05523	3.32803	upregulated	formamidopyrimidine/5-formyluracil/5-hydroxymethyluracil DNA glycosylase
fhuA	P06971	3.27756	upregulated	ferrichrome outer membrane transporter
uxuB	P39160	3.2232	upregulated	D-mannonate oxidoreductase, NAD-binding
cnoX	P77395	3.18488	upregulated	predicted thioredoxin domain-containing protein
flgK	P33235	3.14758	upregulated	flagellar hook-filament junction protein 1
yciW	P76035	2.95402	upregulated	predicted oxidoreductase
zntR	P0ACS5	2.46235	upregulated	DNA-binding transcriptional activator
cheW	P0A964	2.39309	upregulated	purine-binding chemotaxis protein
ribE	P61714	2.33295	upregulated	riboflavin synthase beta chain
bioD	P0A6E9	1.96486	upregulated	predicted dethiobiotin synthetase
dnaJ	P08622	1.64769	upregulated	chaperone Hsp40, co-chaperone with DnaK
hspQ	P0AB20	1.56137	upregulated	DNA-binding protein, hemimethylated
hslR	P0ACG8	1.5463	upregulated	ribosome-associated heat shock protein Hsp15
htpX	P23894	1.50554	upregulated	predicted endopeptidase
pspC	P0AFN2	1.19991	upregulated	transcriptional activator
clpA	P0ABH9	1.12992	upregulated	ATPase and specificity subunit of ClpA-ClpP ATP-dependent serine protease, chaperone activity

bssS	P0AB33	1.05717	upregulated	hypothetical protein
aldB	P37685	-2.6383	downregulated	aldehyde dehydrogenase B
gabT	P22256	-2.0646	downregulated	4-aminobutyrate aminotransferase, PLP-dependent
ydcA	P0ACW4	-1.9356	downregulated	hypothetical protein
prpB	P77541	-1.8059	downregulated	2-methylisocitrate lyase
puuD	P76038	-1.5683	downregulated	gamma-Glu-GABA hydrolase
puuR	P0A9U6	-1.5393	downregulated	DNA-binding transcriptional repressor
gatY	P0C8J6	-1.5302	downregulated	D-tagatose 1,6-bisphosphate aldolase 2, catalytic subunit
ykgF	P77536	-1.4179	downregulated	predicted amino acid dehydrogenase with NAD(P)-binding domain and ferridoxin-like domain
gatZ	P0C8J8	-1.1611	downregulated	D-tagatose 1,6-bisphosphate aldolase 2, subunit

Table 36. Overlap between transcriptomics and proteomics data of differentially expressed genes when comparing cells expressing the scFv in the cytoplasm (cyto) and the periplasm (peri) when induced expression with high concentrations of IPTG (0.125 mM) 24 hours post induction. Genes are upregulated when the scFv is secreted to the periplasm, those that are downregulated decrease when the scFv is secreted to the periplasm. Gene function information is provided by KEGG. Proteomics fold changes are either indicated as upregulated or downregulated as iTRAQ compresses fold change information.

Gene name	Uniprot ID	Log2 fold change transcriptomics	Significant changes proteomics	Function
yciW	P76035	2.55809	upregulated	predicted oxidoreductase
cysH	P17854	2.86299	upregulated	3'-phosphoadenosine 5'-phosphosulfate reductase
fhuA	P06971	2.65417	upregulated	ferrichrome outer membrane transporter
cpxP	P0AE85	1.83413	upregulated	periplasmic protein combats stress
rstA	P52108	1.79247	upregulated	DNA-bindng response regulator in two-component regulatory system with RstB
raiA	P0AD49	1.06387	upregulated	cold shock protein associated with 30S ribosomal subunit
lsrG	P64461	-1.7553	downregulated	conserved hypothetical protein
oppA	P23843	-1.359	downregulated	oligopeptide transporter subunit
xylF	P37387	-1.2822	downregulated	D-xylose transporter subunit

Table 37. Overlap between transcriptomics and proteomics data of differentially expressed genes when comparing cells expressing the scFv in the cytoplasm (cyto) and the periplasm (peri) when induced expression with low concentrations of IPTG (0.0125 mM) 24 hours post induction. Genes are upregulated when the scFv is secreted to the periplasm, those that are downregulated decrease when the scFv is secreted to the periplasm. Gene function information is provided by KEGG. Proteomics fold changes are either indicated as upregulated or downregulated as iTRAQ compresses fold change information.

3.3.3.4.2 Effect of Sec secretion system overloading

3.3.3.4.2.1 Introduction

To gain a better understanding of cellular responses to secretion system overloading, expression of the scFv in the periplasm was compared when adding high (0.125 mM IPTG) and low (0.0125 mM IPTG) concentrations of in. Inducing high levels of scFv expression in the periplasm (0.125 mM) was previously shown to overload the secretion system more than lower levels of induction (0.0125 mM) as less scFv was secreted to the periplasm and more scFv accumulated as inclusion bodies in the cytoplasm (see section 3.3.2.3, 3.3.2.4 and 3.3.2.5). The effect of inducing different levels of the scFv expression in the cytoplasm was also analysed to determine whether responses were unique to secretion system overloading and not just due to the differences in induction levels.

Different levels of scFv expression in the periplasm were analysed 12- and 24-hours post-induction at the transcriptomics level and 24-hours post-induction at the proteomics level. Different levels of scFv expression in the cytoplasm were analysed 24 hours post-induction using both transcriptomics and proteomics (Figure 32).

3.3.3.4.2.2 Transcriptomics overview

Overloading the secretion system by inducing high levels of the scFv expression in the periplasm (0.125 mM IPTG) led to an increase in expression of 546 and 412 genes and decrease in expression of 679 and 647 genes compared to inducing lower levels of scFv secretion (0.0125 mM IPTG), 12- and 24-hours post-induction respectively (Figure 41). Furthermore, there was a large amount of overlap between differentially expressed transcripts 12- or 24-hours post-induction with 298 transcripts being upregulated and 499 transcripts being downregulated in common (Figure 42). This indicates that cellular responses were similar 12 and 24 hours post induction. A full list of differentially expressed transcripts can be found in the supplementary Excel spreadsheet “DGE_Transcriptomics”.

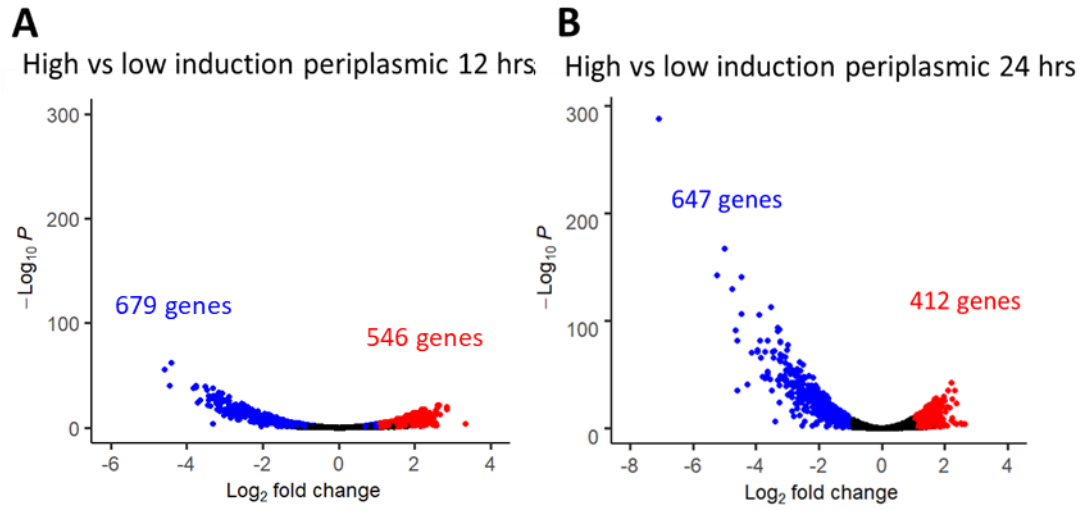


Figure 41. Volcano plots of transcriptomic data comparing the expression of cells secreting the scFv to the periplasm with different levels of inducer (0.125 mM IPTG – secretion system overloading and 0.0125 mM IPTG) 12 hrs (A) and 24 hrs (B) post-induction. Black dots represent genes that are not differentially expressed (adjusted p value > 0.05, log2 fold change between -1 and 1). Red dots represent genes that are upregulated and blue downregulated when inducing high levels of scFv secretion that overload the secretion system. Plots were generated using the EnhancedVolcano R package [248].

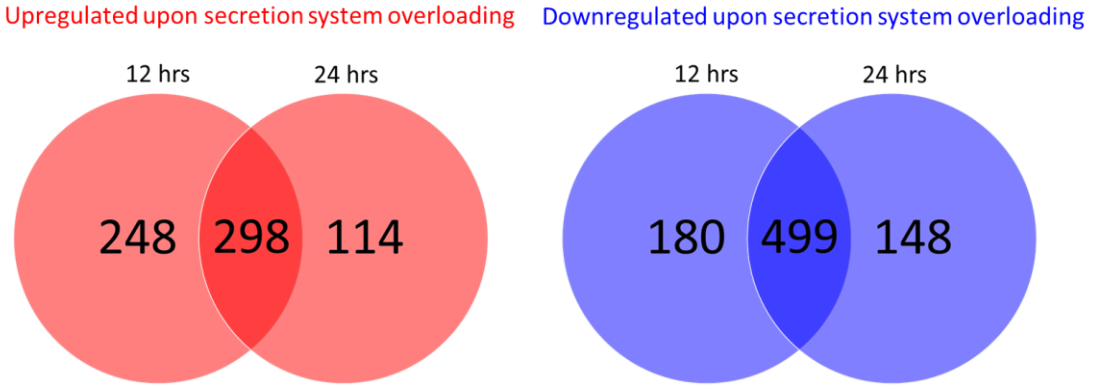


Figure 42. Venn diagram representing overlap when inducing scFv secretion with high (0.125 mM IPTG) and low (0.0125 mM IPTG) induction, 12- and 24-hours post-induction using transcriptomics. Genes are upregulated in samples where the secretion is overloaded (scFv expression with high IPTG) and downregulated when the secretion system is less overloaded (scFv expression with low IPTG).

Transcriptomic data was also obtained comparing *E. coli* expressing the scFv in the cytoplasm with different amounts of inducer (0.125 mM and 0.0125 mM IPTG) 24 hours post-induction (Figure 43). Expression of high levels of the scFv in the cytoplasm led to the upregulation of 30 genes and downregulation of 114 genes, which is considerably less than when expressing different levels of

scFv expression in the periplasm. This was not the focus of this study however it allowed us to determine which responses were unique to secretion and not just due to different induction levels. Indeed, there is an overlap in cellular responses when expressing different levels of the scFv in the cytoplasm and the periplasm with 4 genes being upregulated and 67 genes downregulated in both conditions (Figure 44). This is important to consider when further analysing this data.

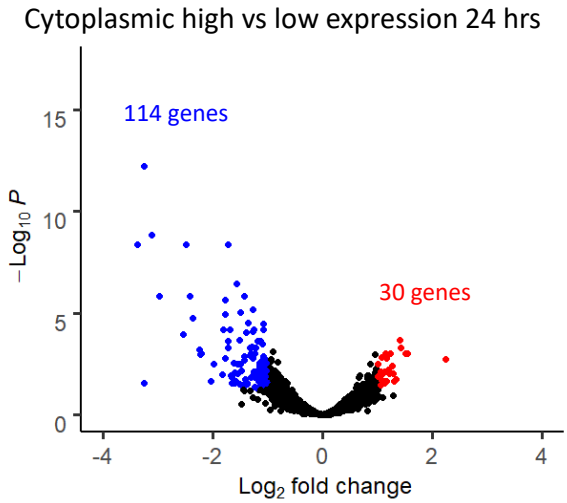


Figure 43. Volcano plots of transcriptomic data comparing the expression of cells secreting the scFv to the cytoplasm with different amounts of inducer (high - 0.125 mM and low 0.0125 mM IPTG) 24 hrs post-induction. A positive fold change indicates genes that were upregulated upon high (0.125 mM) induction and negative fold changes downregulated. Black dots represent genes that are not differentially expressed (adjusted p value > 0.05, log2 fold change between -1 and 1). Red dots represent genes that are upregulated and blue downregulated when inducing high levels of scFv expression in the cytoplasm. Plots were generated using the EnhancedVolcano R package [248].

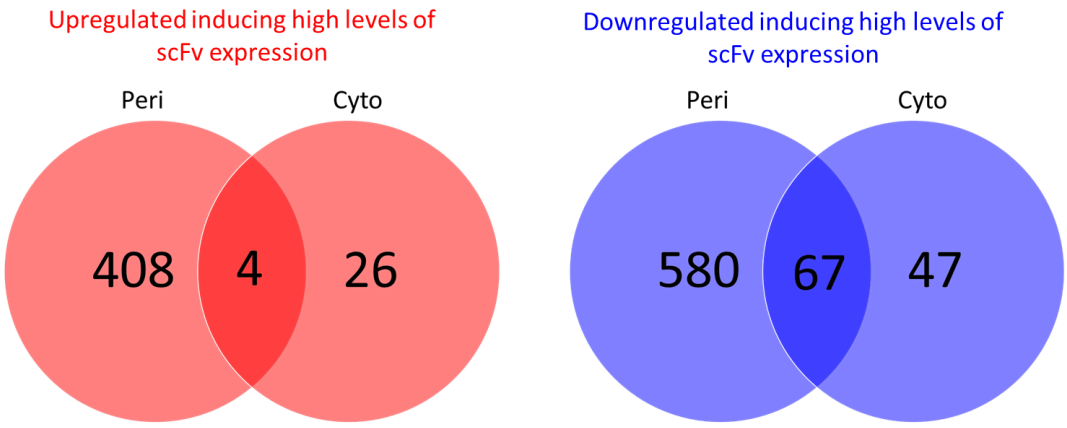


Figure 44. Overlap between transcriptomics data comparing high (0.125 mM) and low induction (0.0125 mM IPTG) when expressing the scFv in the cytoplasm (cyto) and the periplasm (peri) 24 hours post-induction. Genes are upregulated when protein expression is induced with high concentrations of IPTG and downregulated when induced with low concentrations of IPTG.

3.3.3.4.2.3 Proteomics overview

At the proteome scale, 54 proteins were upregulated, and 60 proteins downregulated upon induction of high levels of scFv induction (0.125 mM IPTG) in the periplasm compared to low levels of scFv induction (0.0125 mM IPTG) 24 hours post-induction (Figure 45A). No proteins were identified as differentially expressed when comparing different levels of induction of scFv expression in the cytoplasm (Figure 45B). This could be due to multiple factors including insufficient clustering of these samples (see section 3.3.3.3) and the low number of replicates analysed. It is also worth noting that expressing different levels of the scFv in the cytoplasm had very little effect on gene expression at the transcriptomics level (only 144 transcripts are differentially expressed under these conditions and only 63% of these were quantified using mass spectrometry). Inducing different levels of expression of the scFv in the cytoplasm may therefore have very little impact on the proteome. A full list of differentially expressed proteins can be found in the supplementary Excel spreadsheet "DGE_Proteomics".

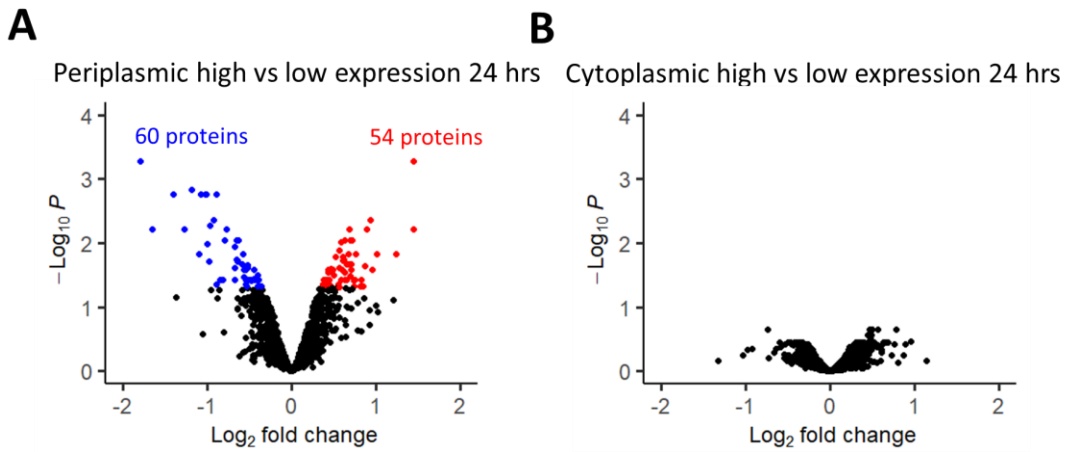


Figure 45. Volcano plots of transcriptomic data comparing the expression of cells secreting the scFv to the periplasm with different levels of inducer (0.125 mM IPTG – secretion system overloading and 0.0125 mM IPTG) 12 hrs (A) and 24 hrs (B) post-induction. Black dots represent genes that are not differentially expressed (adjusted p value > 0.05). No fold change cut-off was used for proteomics data as iTRAQ compresses fold change data. Red dots represent proteins that are upregulated and blue downregulated when inducing high levels of scFv secretion that overload the secretion system. Plots were generated using the EnhancedVolcano R package [248]***.***

3.3.3.4.2.4 Overlap between transcriptomics and proteomics

There was an overlap between transcriptomics and proteomics data comparing high levels and low levels of inducing expression of the scFv in the periplasm 24 hours post-induction, with 15 genes being upregulated and 22 downregulated in common. A list of these genes can be found in Figure 46 which will be discussed in the following sections. Transcript and protein abundance does not necessarily correlate due to differences in stability and therefore if a gene is upregulated at the transcriptomic level this does not necessarily correlate at the proteomic level. It is also important to consider that 98% of the transcriptome was identified using RNA-seq whereas 47% of the proteome was identified using mass spectrometry. Information about whether proteins encoded by differentially expressed transcripts were identified using mass spectrometry can be found in the supplementary Excel spreadsheet "DGE_Transcriptomics"

Gene name	Uniprot ID	Log2 fold change transcriptomics	Significant changes proteomics	Function
yibA	P0ADK6	2.35583	upregulated	lyase containing HEAT-repeat
ibpA	P0C054	2.314393	upregulated	heat shock chaperone
ibpB	P0C058	2.195517	upregulated	heat shock chaperone
zntR	P0ACS5	1.828266	upregulated	DNA-binding transcriptional activator
ribE	P61714	1.683782	upregulated	riboflavin synthase beta chain
flgK	P33235	1.676986	upregulated	flagellar hook-filament junction protein 1
clpX	P0A6H1	1.583283	upregulated	ATPase and specificity subunit of ClpX-ClpP ATP-dependent serine protease
mutM	P05523	1.498162	upregulated	formamidopyrimidine/5-formyluracil/5-hydroxymethyluracil DNA glycosylase
cnoX	P77395	1.493134	upregulated	predicted thioredoxin domain-containing protein
lamb	P02943	1.245586	upregulated	maltose outer membrane porin
clpB	P63284	1.178132	upregulated	protein disaggregation chaperone
hspQ	P0AB20	1.161242	upregulated	DNA-binding protein, hemimethylated
prlC	P27298	1.077011	upregulated	oligopeptidase A
dnaJ	P08622	1.037498	upregulated	chaperone Hsp40, co-chaperone with DnaK
dnaK	P0A6Y8	1.033123	upregulated	chaperone Hsp70, co-chaperone with DnaJ
ymdF	P56614	-7.09068	downregulated	conserved hypothetical protein
prpB	P77541	-3.95954	downregulated	2-methylisocitrate lyase
yebE	P33218	-3.03813	downregulated	conserved hypothetical protein
yqeC	Q46809	-2.81114	downregulated	conserved hypothetical protein
yjdN	P16681	-2.68064	downregulated	conserved hypothetical protein
aldB	P37685	-2.65914	downregulated	aldehyde dehydrogenase B
prpD	P77243	-2.52999	downregulated	2-methylcitrate dehydratase
mgtA	P0ABB8	-2.4201	downregulated	magnesium transporter

rstA	P52108	-2.40097	downregulated	DNA-bindng response regulator in two-component regulatory system with RstB
entB	P0ADI4	-2.18185	downregulated	isochorismatase
yidZ	P31463	-2.1517	downregulated	predicted DNA-binding transcriptional regulator
ygeY	P65807	-2.1292	downregulated	predicted peptidase
yecE	P37348	-2.11208	downregulated	conserved hypothetical protein
cpxP	P0AE85	-2.0336	downregulated	periplasmic protein combats stress
glaH	P76621	-1.88826	downregulated	hypothetical protein
slyD	P0A9K9	-1.48452	downregulated	FKBP-type peptidyl prolyl cis-trans isomerase
katG	P13029	-1.44173	downregulated	catalase/hydroperoxidase HPI(I)
glcC	P0ACL5	-1.41869	downregulated	DNA-binding transcriptional dual regulator, glycolate-binding
oppD	P76027	-1.36149	downregulated	oligopeptide transporter subunit
ydcI	P77171	-1.34026	downregulated	predicted DNA-binding transcriptional regulator
oppF	P77737	-1.23123	downregulated	oligopeptide transporter subunit
pflB	P09373	-1.19514	downregulated	pyruvate formate lyase I

Table 38. Overlap between transcriptomics and proteomics data of differentially expressed genes when comparing different levels of induction (0.125 and 0.0125 mM IPTG) when expressing the scFv in the periplasm 24 hours post-induction. Genes are upregulated increase in expression upon high levels of induction and secretion system overloading, those that are downregulated decrease when the secretion system is more overloaded. Gene function information is provided by KEGG. Proteomics fold changes are only indicated as upregulated or downregulated as iTRAQ is semi quantitative as it compressed fold change data.

3.3.3.4.3 Effect of scFv secretion to the periplasm over time

3.3.3.4.3.1 Introduction

Cellular responses are known to change over time. To analyse how responses changed over time and how cells respond to increasing stress, transcriptomics was used to compare cellular response 12- and 24-hours post-induction when overexpressing the scFv in the periplasm with high (0.125 mM) or low concentrations (0.0125 mM) of IPTG. Proteomics data was not obtained for samples at different timepoints as iTRAQ labels were limited and this was not the main focus of this study.

3.3.3.4.3.2 Transcriptomics overview

59 and 163 genes were upregulated, and 11 and 118 genes were downregulated over time when inducing secretion of the scFv to the periplasm with 0.125 mM IPTG or 0.0125 mM IPTG respectively (Figure 46) indicating that cellular responses were more significant when the secretion system was less overloaded and more scFv was present in the periplasm (0.0125 mM IPTG induction). There was also slight leakiness of samples induced with 0.0125 mM IPTG 24 hours post-induction which could explain this difference in cellular responses over time (see section 3.3.2.5). A full list of differentially expressed transcripts can be found in the supplementary Excel spreadsheet "DGE_Transcriptomics".

There was an overlap between genes that are differentially expressed over time regardless of how much IPTG added to induce scFv expression in the periplasm (0.125 mM or 0.0125 mM) with 41 genes being upregulated and 4 downregulated in common (Figure 47).

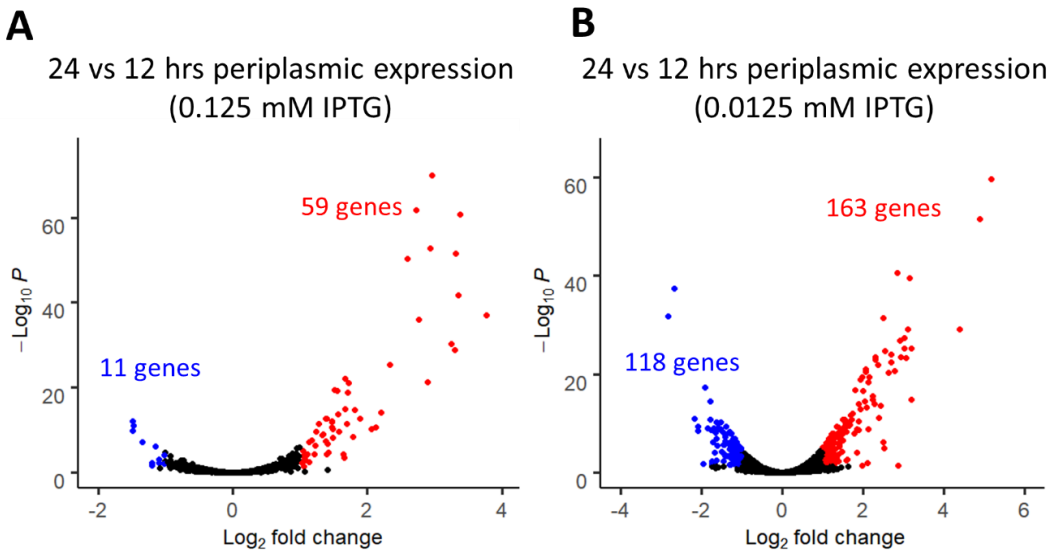


Figure 46. Volcano plots comparing the expression of cells secreting the scFv to the periplasm over time (24 hrs vs 12 hrs post-induction) when inducing with different levels of IPTG (A - 0.125 mM IPTG and B - 0.0125 mM IPTG). Black dots represent genes that are not differentially expressed (adjusted p value > 0.05, log2 fold change between -1 and 1). Red dots represent genes that are upregulated and blue downregulated over time when inducing high levels of scFv secretion that overload the secretion system. Plots were generated using the EnhancedVolcano R package [248].

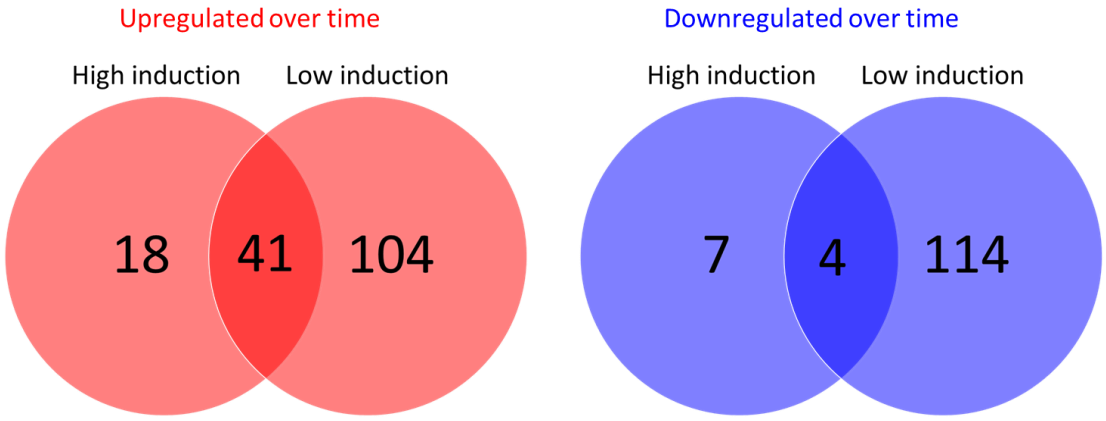


Figure 47. Overlap between transcriptomics data comparing changes in gene expression over time (24 hours vs 12 hours) when expressing the scFv in the periplasm and induced with either a high concentration (0.125 mM) or low concentration (0.0125 mM) IPTG.

3.3.3.5 Functional analysis of differentially expressed genes

Differentially expressed genes were then further analysed to determine their function. Multiple tools were used to analyse gene and protein function, protein localisation, regulatory pathways, and metabolic pathways.

Gene function was analysed using Gene Ontology (GO) enrichment [255] (see example in Figure 48, other enrichment analysis results can be found in Appendix 11, Appendix 12, Appendix 13, Appendix 17, Appendix 18, Appendix 19, Appendix 20, Appendix 24, Appendix 25) and by mapping data to the EcoCyc OMICs dashboard [251], [252] (see example in Figure 49). Gene Ontology enrichment identifies overrepresented functional groups (biological process, molecular function, and cellular component) amongst differentially expressed genes allowing for less biased analysis of large datasets. However, gene ontology enrichment is not suitable for very large or small datasets and Gene Ontology terms are often highly complex and often have issues with annotation making interpretation difficult [264]. The EcoCyc OMICs dashboard groups differentially expressed genes based on their function allowing for easy visualisation of gene and protein function. Furthermore, EcoCyc is the main *E. coli* database and therefore has better gene function annotation and grouping. STEPdb [265] was also used to determine the cellular localisation of proteins including which secretion pathways they use and their disulphide bond content.

Regulatory pathways were analysed using TFInfer [254] (Figure 50) and RegulonDB [266]. TFInfer uses statistical analysis to identify whether transcription factor regulons are differentially expressed. This information was also supplemented with information from the *E. coli* regulatory database RegulonDB which was also used to analyse specific pathways of interest.

Metabolic pathways were analysed using KEGG enrichment to determine whether regulatory pathways were overrepresented and to observe data in a less biased fashion (see example in Figure 51, all other figures can be found in Appendix 14, Appendix 15, Appendix 16, Appendix 21, Appendix 22, Appendix 23). Expression data was also mapped KEGG maps and EcoCyc metabolic maps to observe which sections of metabolic pathways were differentially expressed (see example in Figure 54)

In the following section, this thesis will summarise information from functional analysis of differentially expressed genes and proteins. This has been divided into three sections: stress responses, cellular components, and metabolism.

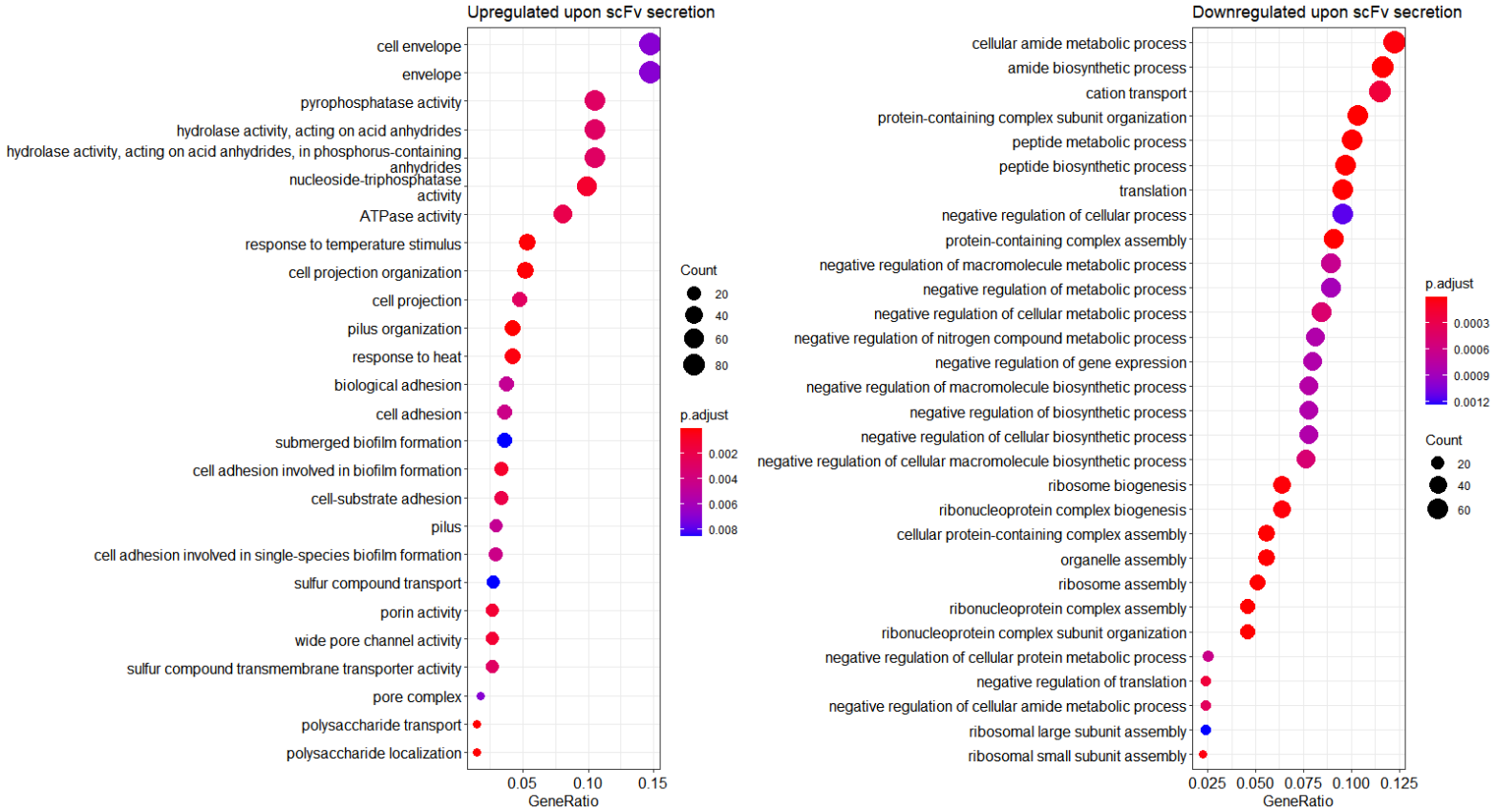


Figure 48. Dot plot representing the 30 most enriched gene ontology terms when using transcriptomics to compare cells overexpressing the scFv in the periplasm and cytoplasm when induced with 0.125 mM IPTG. Enrichment was conducted using molecular function (MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots were conducted using the R package clusterProfiler [255]***.***

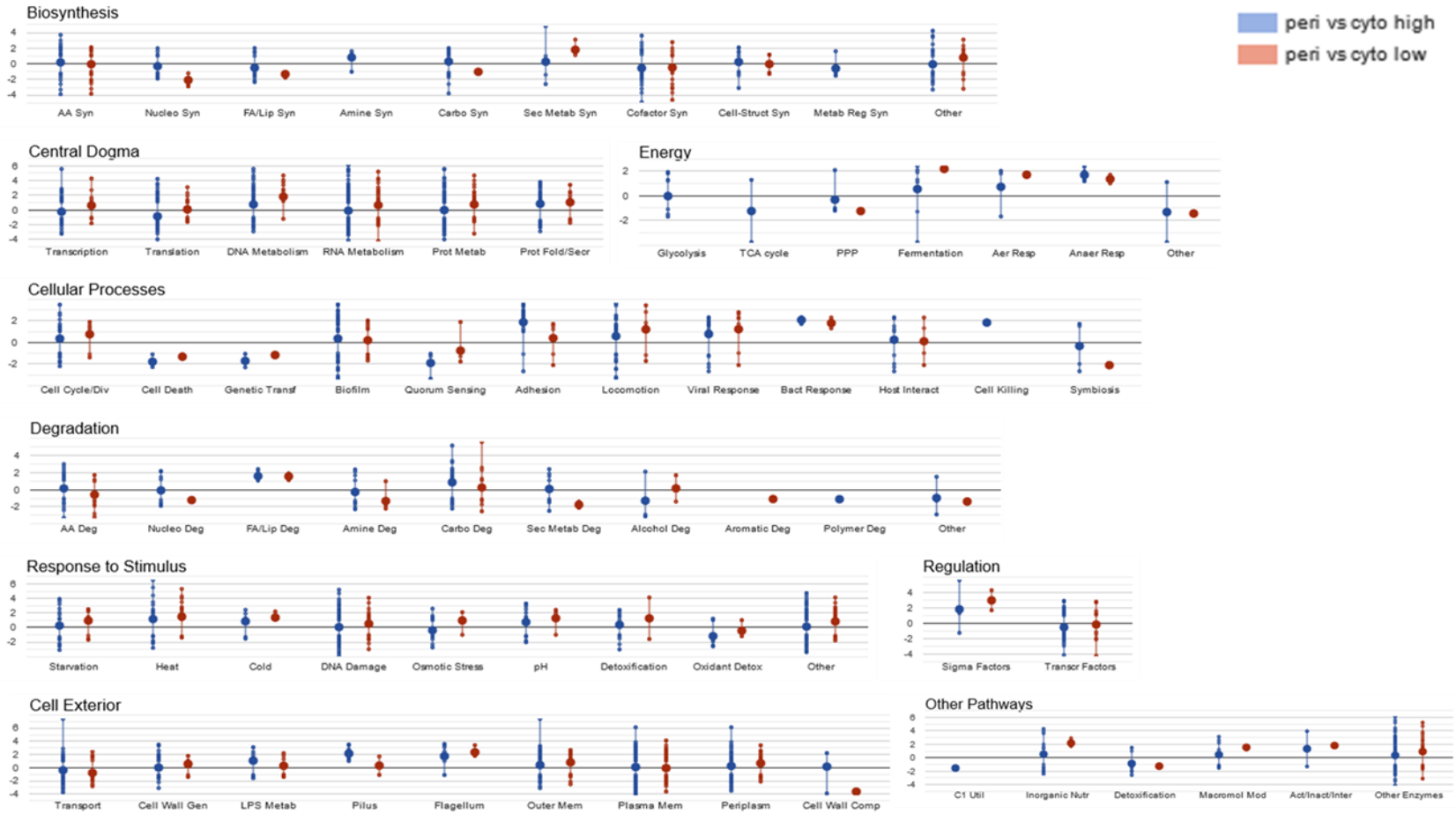


Figure 49. Overview of differentially expressed genes (transcriptomics) when comparing expression of the scFv in the periplasm and cytoplasm when inducing expression with high (0.125 mM blue) and low (0.0125 mM red) concentrations of IPTG. Individual dots represent log2 fold changes for differentially expressed genes, with larger dots representing the average and lines representing the range of values. Figure was created using the EcoCyc Omics dashboard [252]***.***

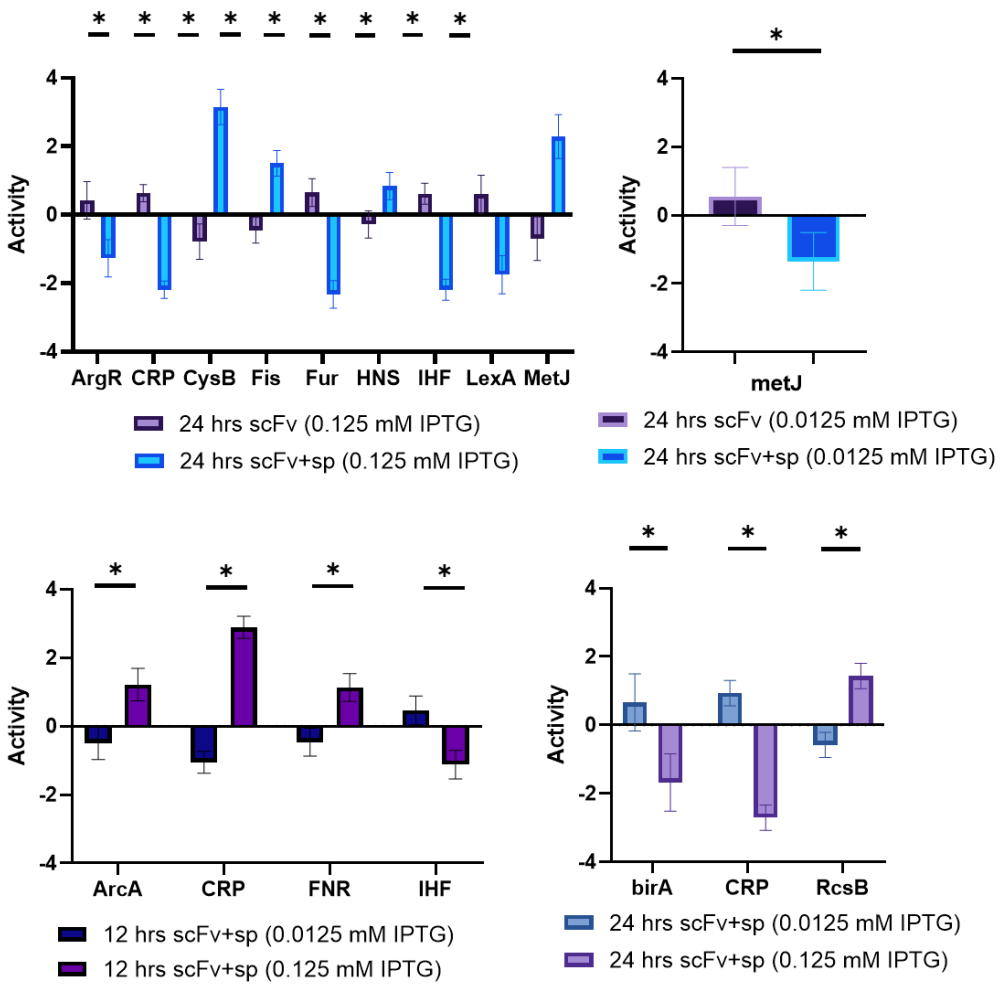


Figure 50. TFinfer analysis indicating statistically significant differences in regulon expression at a transcriptomics level. Samples are either expressing the scFv in the cytoplasm (scFv) or the periplasm (scFv+sp) 12 hrs or 24 hours post-induction and were induced with either high (0.125 mM) or low concentrations (0.0125 mM) IPTG. Regulons were identified as differentially expressed based on one-way ANOVA using GraphPad Prism.

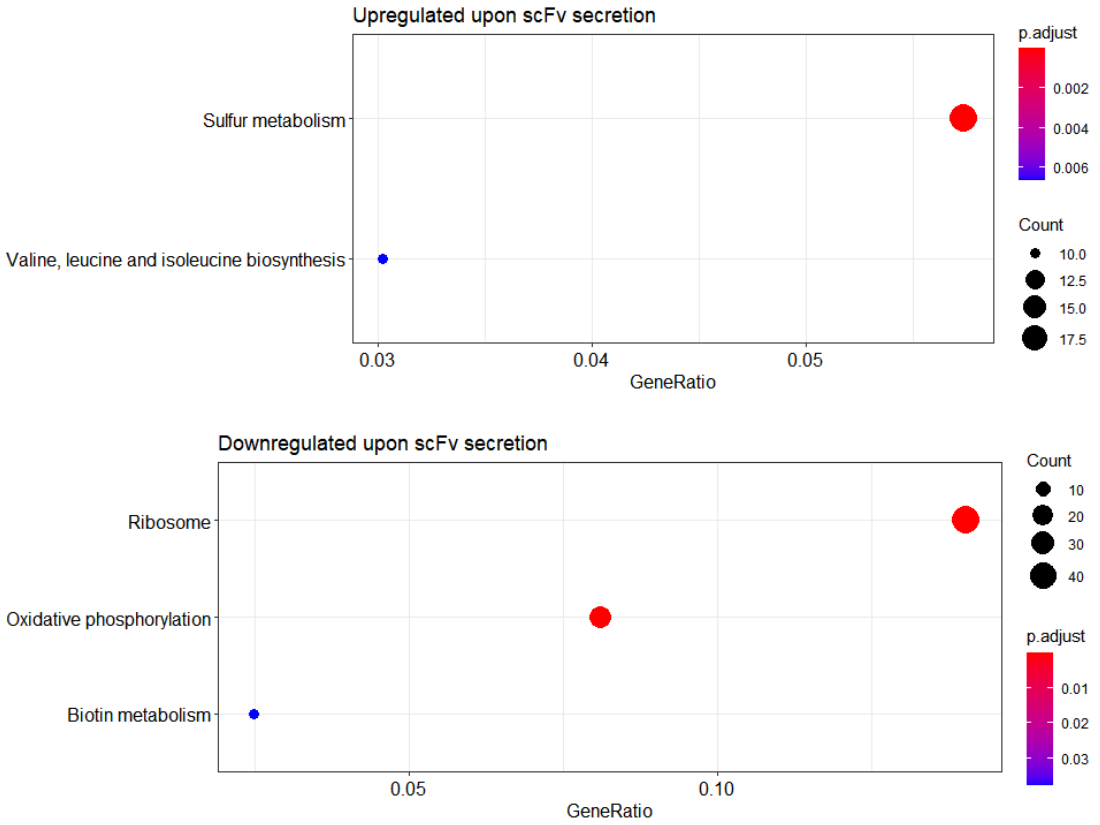


Figure 51. Dot plot representing enriched KEGG terms when using transcriptomics to compare cells overexpressing the scFv in the periplasm and cytoplasm while induced with 0.125 mM IPTG. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255]. ***The size of each dot represents the number of differentially expressed genes in said KEGG pathway and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes within the KEGG pathway. Pathways are considered enriched if they have an adjusted p-value < 0.05.***

3.3.3.5.1 Stress responses

3.3.3.5.1.1 Envelope stress response

The envelope stress response was of particular interest in this study as it is activated by the accumulation of unfolded periplasmic and membrane proteins and helps relieve stress caused by membrane damage [65], [66], [78]–[81] (see section 1.1.2.1.4). The envelope stress response pathway has multiple regulators (CpxR, PspF, SigmaE (RpoE), BaeR and RcsB).

ScFv secretion led to an increase in expression of the majority of the *psp* operon that further increased over time based on transcriptomics analysis. This increase was more intense when lower amounts of IPTG were used to induce scFv expression that overloaded the Sec secretion system less (Figure 52). This is consistent with proteomics data that indicated that high levels of scFv expression in the periplasm led to an increase in PspC expression. Previous studies have also observed that recombinant protein secretion activated the psp response [146], [207]. Interestingly the PspF regulon was not identified as differentially expressed based on statistical analysis (TFinfer). This may be due to multiple factors including improper annotation of regulons or due to the fact that few genes are regulated by PspF and a large number of transcripts were identified as differentially epxressed.

There was a large amount of variation in the *rpoE* and *cpx* response which makes this data difficult to interpret (Figure 52). However, on average there was an increase in the *rpoE* response when secreting the scFv to the periplasm regardless of the amount of inducer added indicating that the RpoE responds to scFv secretion regardless of the level of secretion system overloading (Figure 52). The *cpx* response on the other hand only increased when inducing low levels of scFv secretion that did not overload the secretion system (Figure 52). There was also an increase in expression of the negative regulator of the *cpx* response pathway CpxP at the proteomics level upon high induction. This was not expected as the *cpx* response is known to respond to unfolding of membrane proteins which is known to occur when the secretion system is overloaded [81], [82], [97], [267]–[270]. However, this may just be due to a large amount of

overlap between the *cpx* response and other stress response pathways [65], [81], [271].

Two genes regulated by the *bae* response (*spy* and *mtdT*) and genes regulated by the RcsB response were downregulated upon scFv secretion indicating that these envelope stress response pathways were not activated upon scFv secretion (Figure 52).

The envelope stress response regulates expression of multiple genes involved in membrane and periplasmic protein folding [81], [82], [270], [272]. These will be covered in more depth in section 3.3.3.5.2.2.

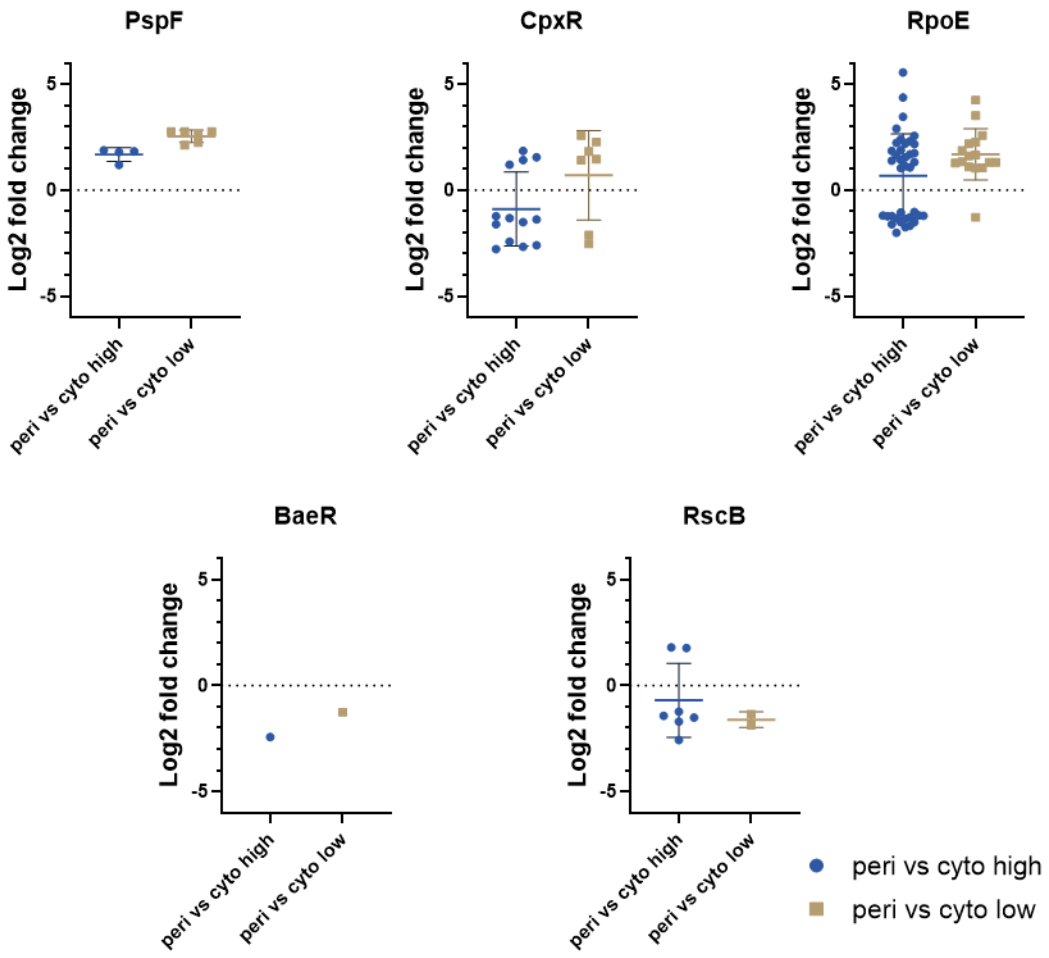


Figure 52. Figure representing differentially expressed transcripts that are regulated by the envelope stress response (CpxR, PspF, RpoE, BaeR and RscB) when comparing strains expressing the scFv in the cytoplasm (cyto) and the periplasm (peri) when induced with high (0.125 mM) and low (0.0125 mM) concentrations of IPTG. Lines represent average Log2 fold change and error bars represent 1 standard deviation. Regulons were determined using RegulonDB [266] ***and data were transformed based on information about whether gene expression was positively or negatively affected (e.g. if a gene showed a -1 fold difference and was negatively regulated by CpxR this would be transformed to a 1 fold difference).***

3.3.3.5.1.2 Heat shock response

The heat shock response is similar to the envelope stress response but responds to unfolded cytoplasmic proteins. There was an increase in the expression of genes involved in the heat shock response upon secretion of the scFv in the periplasm, particularly when inducing high levels of scFv expression in the periplasm, which increased over time. This was confirmed by gene ontology enrichment analysis (Figure 48, Appendix 11, Appendix 12, Appendix

18, Appendix 20, Figure 53) and the increase in expression of the RpoH protein, the main regulator of the heat shock response when inducing high levels of scFv secretion (0.125 mM IPTG). This activation of the heat shock response could be explained by more of the scFv remaining insoluble in the cytoplasm under these conditions based on SDS-PAGE analysis conducted previously (see section 3.3.2.6). The heat shock response regulates expression of cytoplasmic protein folding and degradation genes which will be covered in more depth in the following sections (see 3.3.3.5.2.2). This is consistent with other studies that have found that overexpression of recombinant proteins leads to an increase in expression of the heat shock response, particularly when proteins form insoluble inclusion bodies [168], [205], [212], [229], [273].

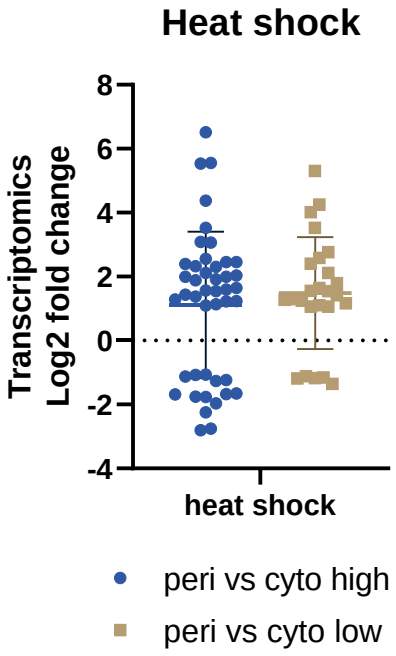


Figure 53. Differentially expressed transcripts involved in the heat shock response (GO 0009408), when comparing strains expressing the scFv in the periplasm (peri) and the cytoplasm (cyto) when induced with high (0.125 mM) and low (0.0125 mM) concentrations of IPTG. Figure represents the average fold change and error bars one standard deviation.

3.3.3.5.1.3 Iron stress

ScFv secretion led to an increase in the expression of genes involved in iron starvation which increased over time. This increase in iron starvation was

more intense when inducing low levels of scFv secretion that overloaded the secretion system less.

This was confirmed by TFinfer analysis that indicated that genes regulated Fur which negatively regulates iron uptake were downregulated upon scFv secretion in transcriptomics experiments (Figure 50). Enrichment analysis also identified that genes involved in iron transport, siderophore synthesis and iron sulphur cluster assembly were upregulated at the transcriptomics scale (Figure 54, Figure 48, Appendix 12, Appendix 17, Appendix 18, Appendix 23, Appendix 24). Furthermore, the iron transporter FhuA was also upregulated at the proteomics scale upon scFv secretion.

Iron starvation responses are known the be activated by the envelope stress response as many important membrane proteins including the electron transport chain, contain iron sulphur clusters [82], [84], [274]. Previous transcriptomics studies also found that overexpressing recombinant proteins in the cytoplasm led to an increase in the expression of genes involved in iron homeostasis and the Fur regulon [211], indicating that activation of iron stress is not unique to expression of recombinant proteins in the periplasm.

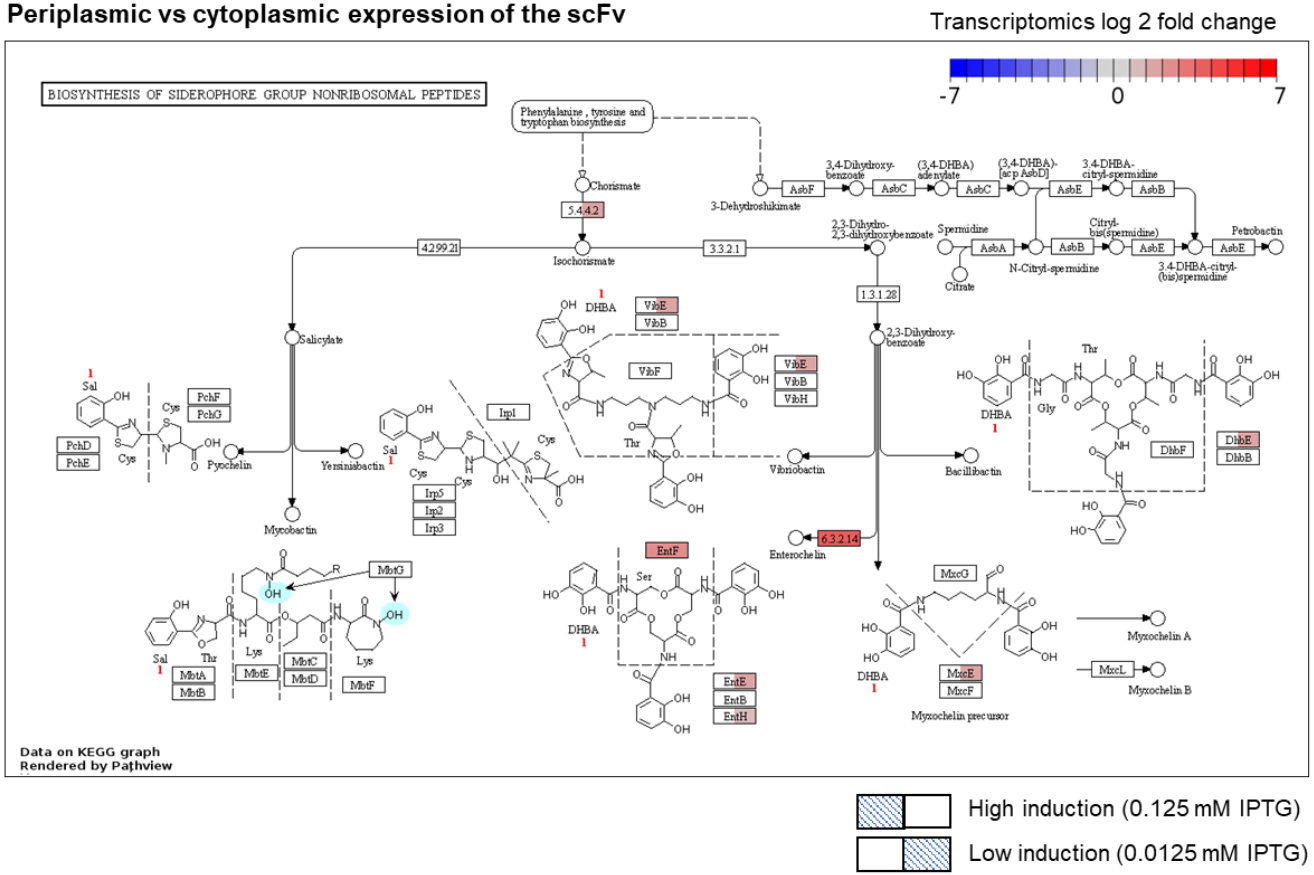


Figure 54. KEGG map of siderophore synthesis pathways (ecj01053) containing differentially expressed genes when comparing expression of the scFv in the periplasm and cytoplasm when induced with a high (0.125 mM left) and low (0.0125 mM right) concentrations of IPTG. Colours represent log2 fold change of transcriptomics data with genes upregulated upon secretion in red and downregulated upon secretion in blue. Data was visualised using Pathview [275], [276]***.***

3.3.3.5.1.4 Oxidative stress

High induction of scFv secretion (secretion system overloading) led to a decrease in expression of genes involved in oxidative stress compared to lower levels of induction as identified through enrichment analysis (Figure 55, "response to oxygen containing compound" Appendix 18). The oxidative stress response is known to be linked to iron homeostasis (discussed in section 3.3.3.5.1.3). Oxidative stress may be linked to increased levels of scFv in the periplasm when inducing low levels of scFv expression as this is known to affect membrane integrity.

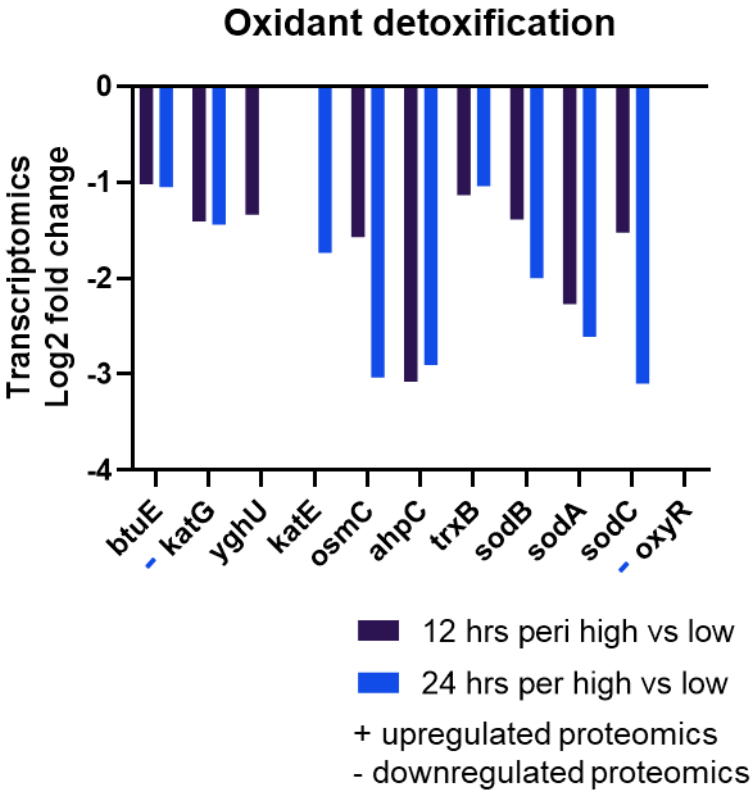


Figure 55. Differentially expressed genes and proteins involved in oxidant detoxification stress (GO 0098869) when comparing cells expressing high (0.125 mM IPTG) and low (0.0125 mM IPTG) levels of periplasmic scFv (peri) 12 hrs (purple) and 24 hours (blue) post-induction. Transcriptomic log2 fold changes are indicated as well as upregulated proteins (plus sign) and downregulated proteins (minus sign). Proteomics data is only available 24 hours post-induction.

3.3.3.5.1.5 SOS response and DNA damage

ScFv secretion also activated the SOS response. Genes regulated by LexA were downregulated upon scFv secretion based on TFinfer analysis (Figure 50).

LexA is a repressor of the SOS response and controls expression of genes involved in DNA damage. Other studies have also found that recombinant protein overexpression led to an increase in the SOS response [277], however, this increase was not consistent between studies indicating that multiple factors other than recombinant protein overexpression may be influencing activation of the SOS response [157].

3.3.3.5.1.6 Osmotic stress

Multiple transcripts involved in the osmotic stress response (*osmB, osmE* and *osmY*) were downregulated over time (24 vs 12 hours post-induction) when inducing low levels of scFv expression (Figure 56). Inducing a low amount of scFv secretion to the periplasm leads to increased scFv yields in the periplasm, and increased cell leakiness which may affect membrane stability over time.

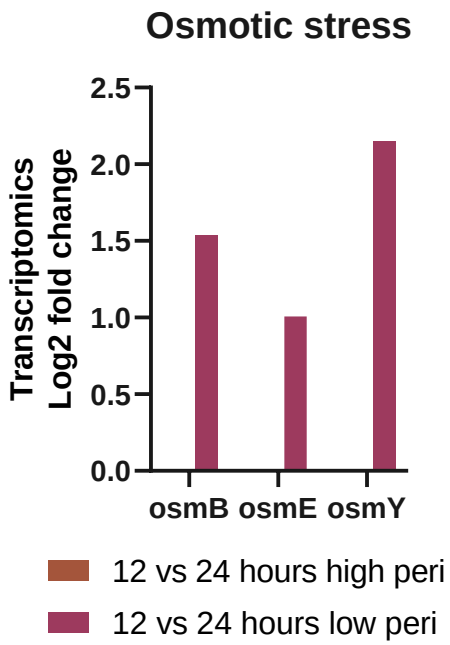


Figure 56. Differentially expressed genes involved in osmotic stress (GO 0006970) over time when expressing the scFv in the periplasm and adding either high (0.125 mM IPTG orange) or low (0,0125 mM red) concentrations of IPTG.

3.3.3.5.1.7 Other stress response pathways

TFinfer analysis also identified several broad stress responses that were differentially expressed upon scFv secretion (Figure 50).

ScFv secretion activated expression of exponential response pathways that promote synthesis of ribosomal RNA and genes involved in amino acid synthesis. The regulator RpoD was upregulated at both the transcript (both IPTG concentrations) and protein level (high induction) upon scFv secretion. The Fis regulon was also upregulated upon scFv secretion regardless of the amount of inducer added based on TFInfer analysis (Figure 50). This is consistent with strains maintaining growth under these conditions and is in line with an increase in transcription and translation upon scFv overexpression.

The general stress response regulates expression of many stress pathways. High levels of scFv secretion led to a decrease in expression of the general stress response regulator *rpoS* (transcriptomics but not proteomics). Expression of several general stress response regulons were also downregulated upon scFv secretion based on TFinfer analysis including a decrease in the CRP and IHF regulon (particularly when inducing high levels of scFv expression and overloading the secretion system) and an increase in the HNS (a repressor). This indicates that scFv secretion does not activate the general stress response pathway.

3.3.3.5.2 Cellular components

3.3.3.5.2.1 Protein secretion

ScFv secretion to the periplasm affected expression of several genes involved in protein secretion (Figure 57).

Expression of the Sec secretion system is of particular interest to this study as it is used to secrete our protein of interest (the scFv). ScFv secretion led to a decrease in expression of transcripts that encode the Sec translocon *secYE*, the targeting chaperone *secB* and membrane protein insertases/chaperones *yidC*/*yajC* and an increase in expression of the signal peptidase *lepB* (Figure 57). This response was more prominent when inducing high levels of scFv expression that overloaded the secretion system. A decrease in expression of the Sec secretion system could potentially be detrimental to secretion of recombinant proteins as well as native membrane and periplasmic proteins however it is important to note that this change was not seen at the protein level. This opposes

findings from a proteomics study by others that found that expressing HGH in the periplasm led to an increase in expression of the Sec-translocon protein SecA, the signal peptidase LepB, and the membrane protein integrase/chaperone YidC, which further decreased when removing the inducer [106]. This difference between studies indicates that this increase in expression of Sec components may be specific to culture/induction conditions, properties of the protein overexpressed and levels of secretion system overloading.

ScFv secretion also led to a decrease in expression of genes involved in the Tat secretion system pathway (*tatAE*) at the transcriptional level which was more prominent when inducing high levels of scFv secretion and overloading the Sec secretion system (Figure 57). Even though scFv secretion was not dependent on this pathway, the Tat secretion system is the other main secretion system used to export periplasmic proteins and therefore a decrease in its expression could also be detrimental to the cell. This is consistent with a proteomics study that found that secreting recombinant proteins via the Tat secretion pathway had a negative impact on Tat proteins [44].

ScFv secretion led to an increase in expression of several affected expression of type II general secretory pathway genes, including inner membrane proteins *gspCGHIKM* and the ATPase *gspE* (Figure 57). This increase was only seen when inducing high levels of scFv secretion that overloaded the secretion system and was only determined at the transcriptomic level as no Gsp proteins were identified using mass spectrometry. Gsp proteins are secreted by the Sec secretory pathway and are used to secrete toxins indicating that recombinant protein secretion may potentially be causing increased levels of stress.

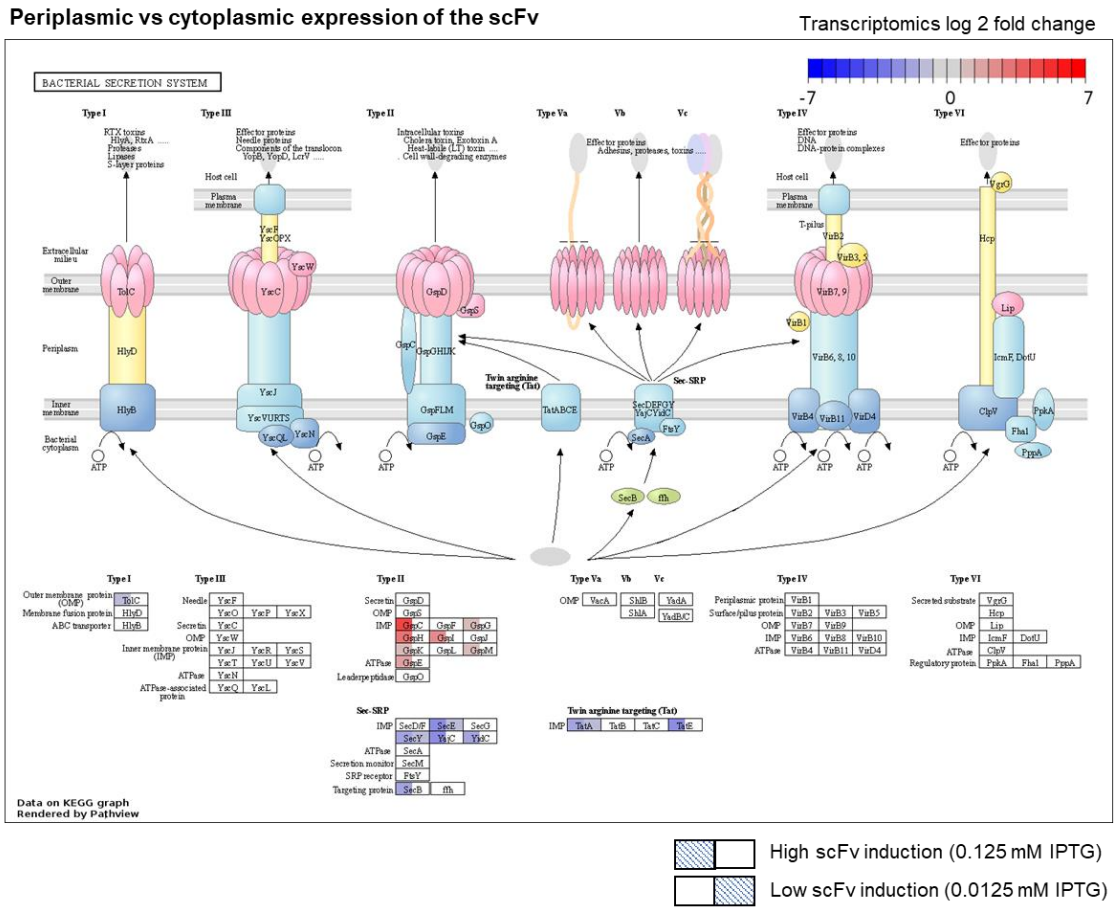


Figure 57. KEGG map of bacterial secretion systems (ecj03070) containing differentially expressed genes when comparing expression of the scFv in the periplasm and cytoplasm. Samples were generated when inducing either high (left) or low (right) levels of scFv expression. Colours represent log2 fold change of transcriptomics data with genes upregulated upon secretion in red and downregulated upon secretion in blue. Data was visualised using Pathview [275], [276]***.***

3.3.3.5.2.2 Protein folding/degradation

ScFv secretion affected expression of several genes involved in protein folding and degradation as identified through enrichment analysis ("protein folding", "unfolded protein binding" "regulation of protein stability") (Appendix 11, Appendix 18, Appendix 20, Figure 58, Figure 59). These proteins can be divided into three categories: proteins that aid folding in the cytoplasm, periplasm and membrane.

There was an increase in expression of several cytoplasmic chaperones and proteases upon scFv secretion (Figure 58, Figure 59). This is consistent with previous observations that the heat shock response is activated upon scFv

secretion (see section 3.3.3.5.1.2). This increase was more intense when inducing high levels of scFv secretion that overloaded the secretion system (Figure 58, Figure 59) and is consistent with more of the scFv forming inclusion bodies in the cytoplasm (see section 3.3.2.6). Indeed, multiple other studies have found that recombinant protein overexpression activates expression of chaperones and proteases [168], [205], [212], [229], [273]. Activation of cytoplasmic chaperones and proteases can be beneficial to recombinant protein secretion as it prevents proteins from folding in the cytoplasm, increases degradation of incorrectly folded proteins and alleviates stress associated with inclusion bodies [63]. However, expression of certain chaperones such as Tig can be detrimental as they sequester secreted proteins [63], [130], [131].

On the other hand, a large number of genes involved in periplasmic folding were downregulated (*dsbB, ivy, spy, oppA, hdeA, dppA, osmY and ppiA),* particularly when inducing high levels of scFv expression that overloaded the secretion system (Figure 58). This decrease in expression of genes involved in periplasmic protein folding could potentially be detrimental to product yields although it is important to note that none of these changes were seen at the protein level. These differences in expression of periplasmic protein folding genes could be explained by changes in the envelope stress response which is discussed in section 3.3.3.5.1.1.

ScFv secretion also led to an increase in expression of several periplasmic and membrane proteases. Many of these genes are regulated by the envelope stress response which was previously shown to be altered upon scFv expression (discussed in section 3.3.3.5.1.1) (Figure 59). It was particularly interesting that scFv secretion affected expression of membrane proteases (FtsH, HtpX) and their negative regulators (HflCK, YccA), with an increase in expression of HtpX being seen at both the transcriptome and proteomics level as these proteases play a role in degrading SecY which forms part of the Sec secretion system used to secrete the scFv [278]–[281].

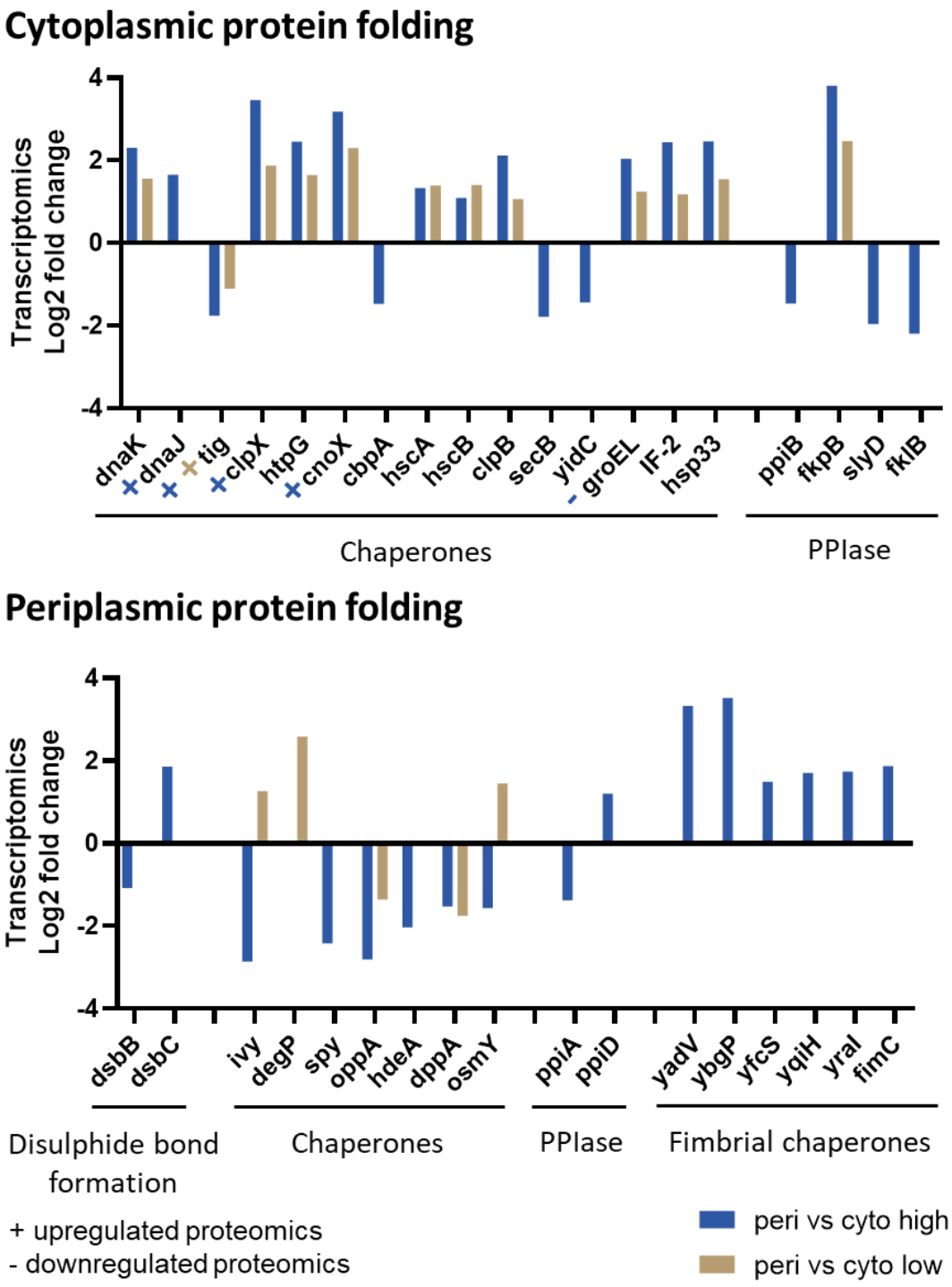


Figure 58. Differentially expressed proteins and transcripts involved in protein folding (GO 0006457), when comparing strains expressing the scFv in the periplasm (peri) and the cytoplasm (cyto) when induced with high (0.125 mM) and low (0.0125 mM) concentrations of IPTG. Cellular location of proteins was determined using STEPdb [265] ***and genes were further divided based on their function. Log2 transcript fold change is indicated as well as where a protein was upregulated (plus sign) or downregulated (minus sign).***

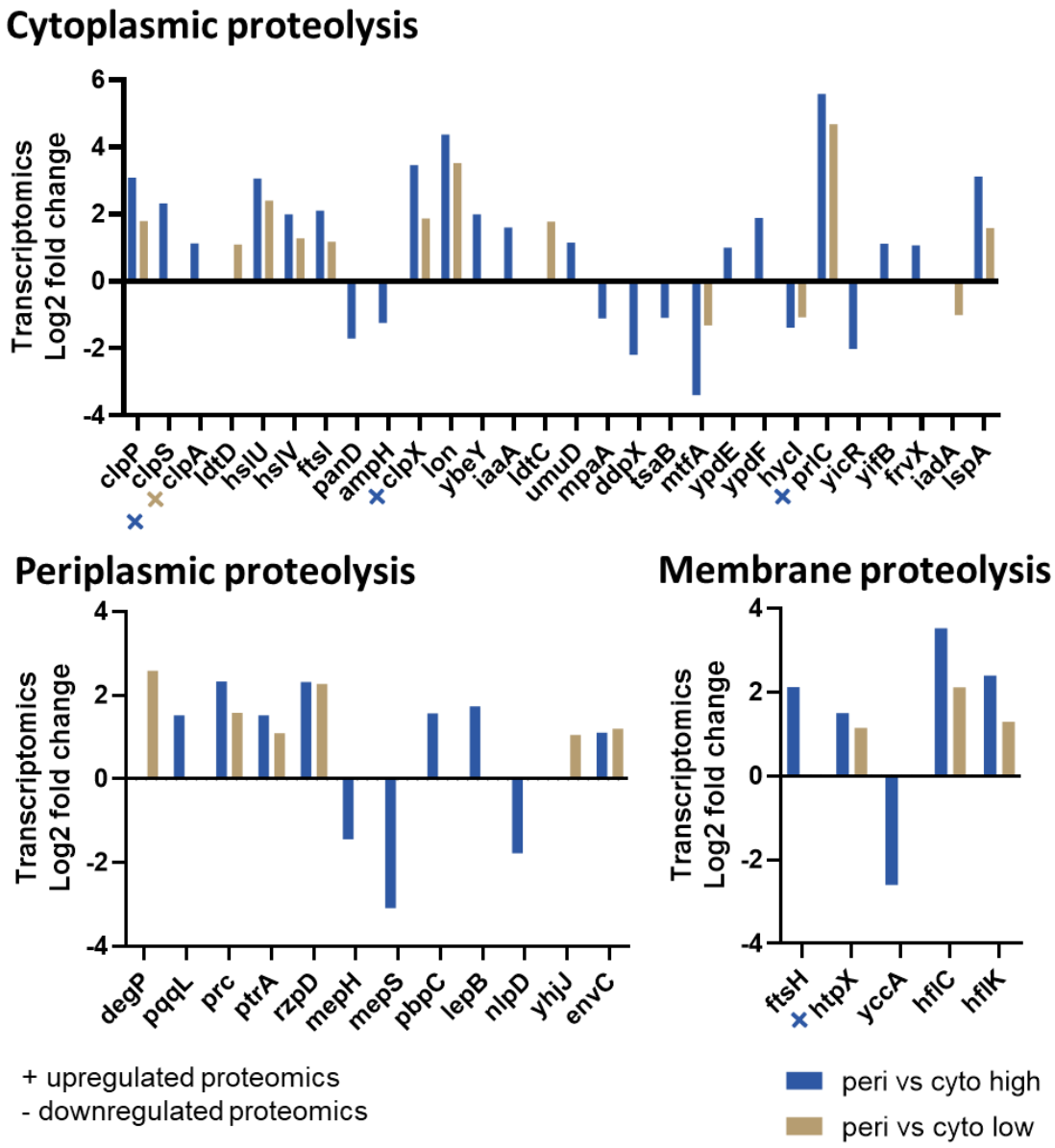


Figure 59. Differentially expressed transcripts involved in protein degradation (GO 0006508), when comparing strains expressing the scFv in the periplasm (peri) and the cytoplasm (cyto) when induced with high (0.125 mM blue) and low (0.0125 mM yellow) concentrations of IPTG. Cellular location of proteins was determined using STEPdb [265]***. YccA and hflCK (negative regulators of ftsH) were also added to the list of genes involved in proteolysis.***

3.3.3.5.2.3 Pilli/fimbriae

Secretion of the scFv to the periplasm led to an increase in expression of well characterised and putative pilus/fimbriae proteins when inducing high levels of scFv secretion that overloaded the secretion system (Figure 60). Indeed "pilus", "pilus organisation" and "pilus assembly" gene ontology terms were enriched amongst these samples (Figure 48, Appendix 17, Appendix 18).

Furthermore, high levels of scFv secretion also led to an increase in expression of pilus chaperones (*yadV, ybgP. yfcS, yaiH, yraI* and *fimC*) as seen in section 3.3.3.5.2.2. This increase in expression of pilus proteins was only seen at the transcript level however upon further observation this was because no pilus proteins were detected using mass spectrometry. This increase could potentially be linked to the activation of envelope stress response pathways (covered in section 3.3.3.5.1.1) as this stress response is known to be activated by misfolded pilus proteins [269], [271], [282], [283].

Other omics studies have also found that overexpressing recombinant proteins in the cytoplasm led to an increase in expression of pilus proteins [202]. This indicates that an increase in expression of pilus proteins can also be due to expression of recombinant proteins in the cytoplasm.

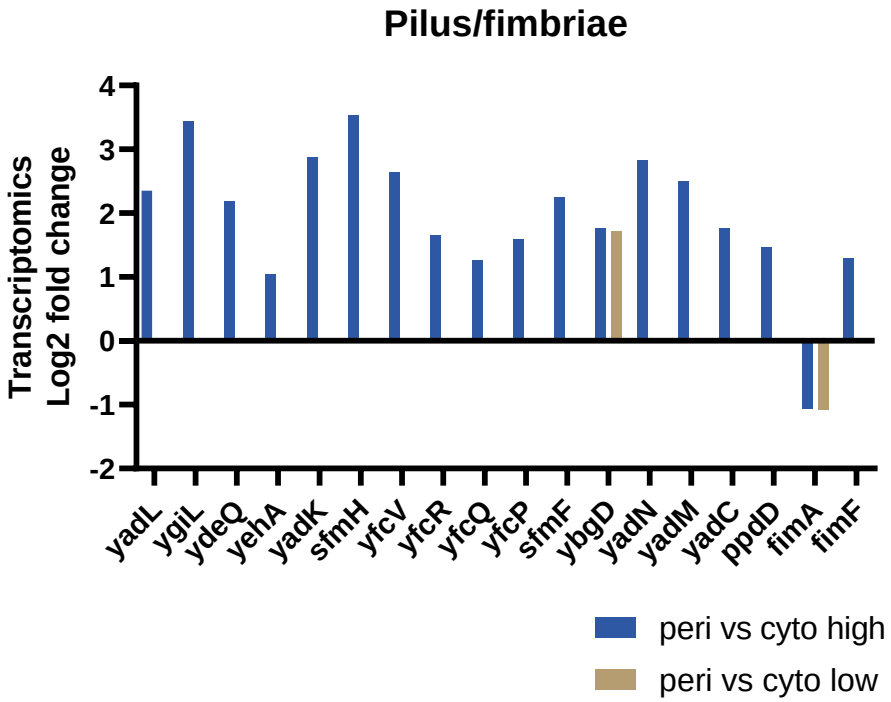


Figure 60. Differentially expressed pilus genes (GO 0009289), when comparing strains expressing the scFv in the periplasm (peri) and the cytoplasm (cyto) when induced with high (0.125 mM blue) and low (0.0125 mM yellow) concentrations of IPTG.

3.3.3.5.2.4 Flagella

Secretion of the scFv to the periplasm led to an increase in expression of genes involved in flagella synthesis, particularly when inducing high levels of

scFv secretion and overloading the secretion system at both the protein and transcript level as identified through KEGG enrichment (Appendix 22). This included an increase in flagella basal bodies, hook and regulatory proteins (flgM, fliA) (Figure 61).

This is consistent with other OMICs studies have also found that overexpressing recombinant proteins in the cytoplasm or inner membrane led to an increase in expression of flagella synthesis genes [202], [204], [209]. Furthermore, one study also found that overexpression of inner membrane proteins led to an increase in the length of flagella [209].

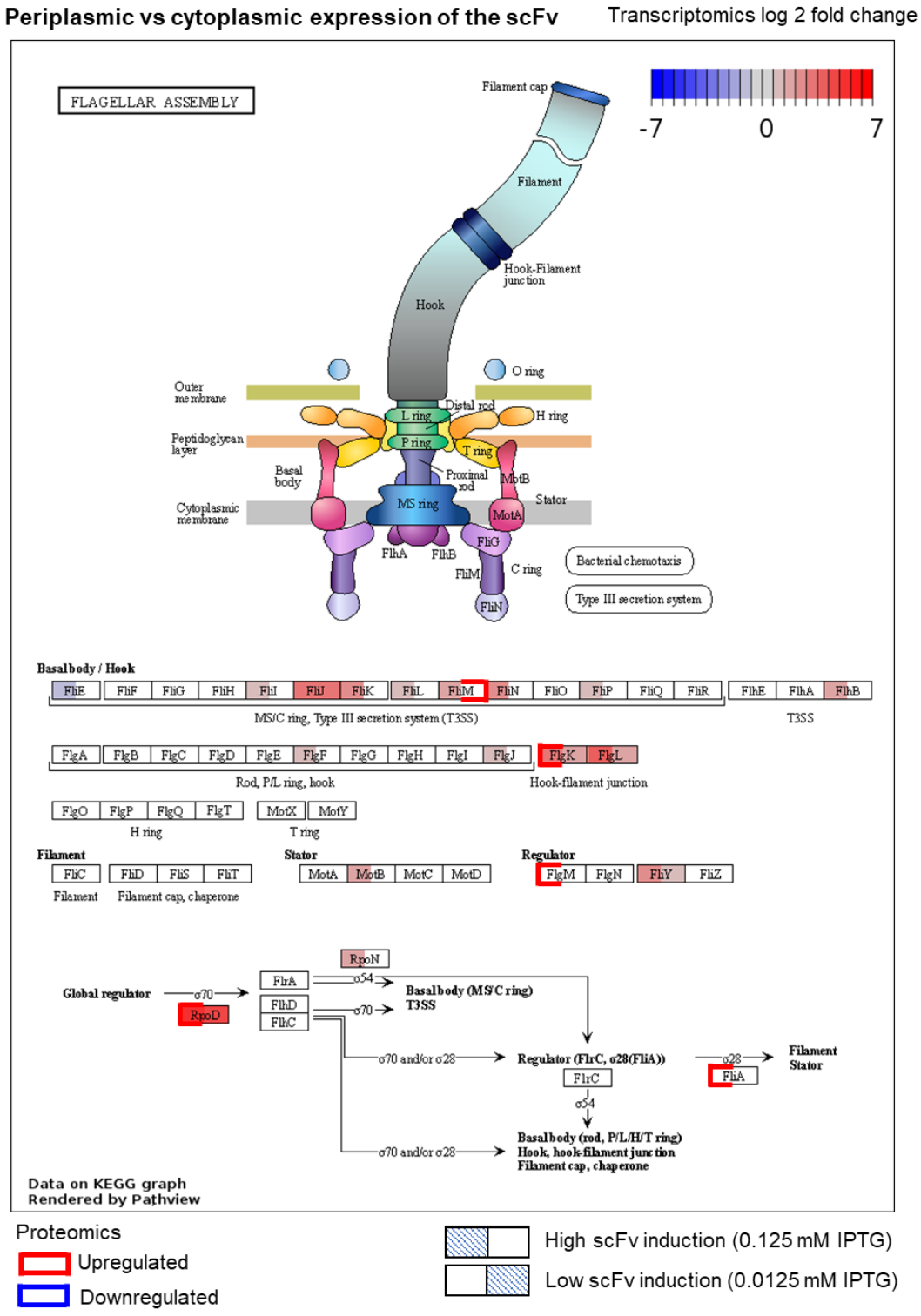


Figure 61. KEGG map of flagella synthesis pathway (ecj02040) containing differentially expressed genes and proteins when comparing expression of the scFv in the periplasm and cytoplasm. Samples were generated when inducing either high (left) or low (right) levels of scFv expression. Differentially expressed proteins are circled in red if they are upregulated and blue if they are downregulated. Transcriptomic log2 fold changes of differentially expressed transcripts are coloured in red if upregulated and blue if downregulated. Data was visualised using Pathview [275], [276].

3.3.3.5.2.5 Envelope proteins

Inducing high levels of secretion of the scFv to the periplasm led to a decrease in expression of periplasmic and membrane proteins as identified through enrichment analysis (Appendix 13). Indeed 7 out of 9 proteins that were downregulated upon high levels of scFv secretion are secreted via the Sec system (XylF, ThiB, OppA, YdcA, PotF, GltI, PspE). This decrease was however not seen at the transcript level or when inducing lower levels of scFv secretion that overloaded the secretion system less. This could be due to secretion system overloading affecting the secretion of native *E. coli* proteins, their folding and therefore their abundance. Indeed previous proteomics analysis identified that native membrane and periplasmic proteins were present in inclusion bodies upon secretion system overloading indicating that they had not been properly secreted, however, this could also be due to improper extraction of inclusion bodies [219], [284]. Other studies however have not identified a decrease in expression of envelope proteins when secreting recombinant proteins to the periplasm [90].

3.3.3.5.3 Metabolism

3.3.3.5.3.1 Central metabolism

ScFv secretion led to a decrease in expression of several genes involved in oxidative phosphorylation/the electron transport chain which was more severe when inducing high levels of expression that overloaded the secretion system and was more severe over time (Figure 62). This decrease was identified through enrichment analysis of transcriptomics data (Figure 51, Appendix 14, Appendix 17, Appendix 18, Appendix 21, Appendix 22) and affected expression of genes that encoded the NADH dehydrogenase (the *nuo* operon), succinate dehydrogenase (the *sdh* operon), cytochrome c oxidase (*cyo* operon), cytochrome bd (*cydAB*) as well as the ATP synthase. This decrease was however not seen at the proteomics level, and on the contrary, there was an increase in expression of the beta subunit of the ATPase upon high levels of scFv expression (Figure 62). The electron transport chain/oxidative phosphorylation pathway is a major player in ATP synthesis and the generation proton motive force which

plays an important role in secretion (see introduction section 1.1.2.1.2). The oxidative phosphorylation pathway is negatively regulated by the envelope stress response pathway which was discussed previously [283]. These observations are consistent with previous studies that found that overexpression of periplasmic and membrane proteins led to a decrease in expression of genes involved in oxidative phosphorylation [219], [273] indicating that this response may not be unique to secretion of our anti-TNF scFv.

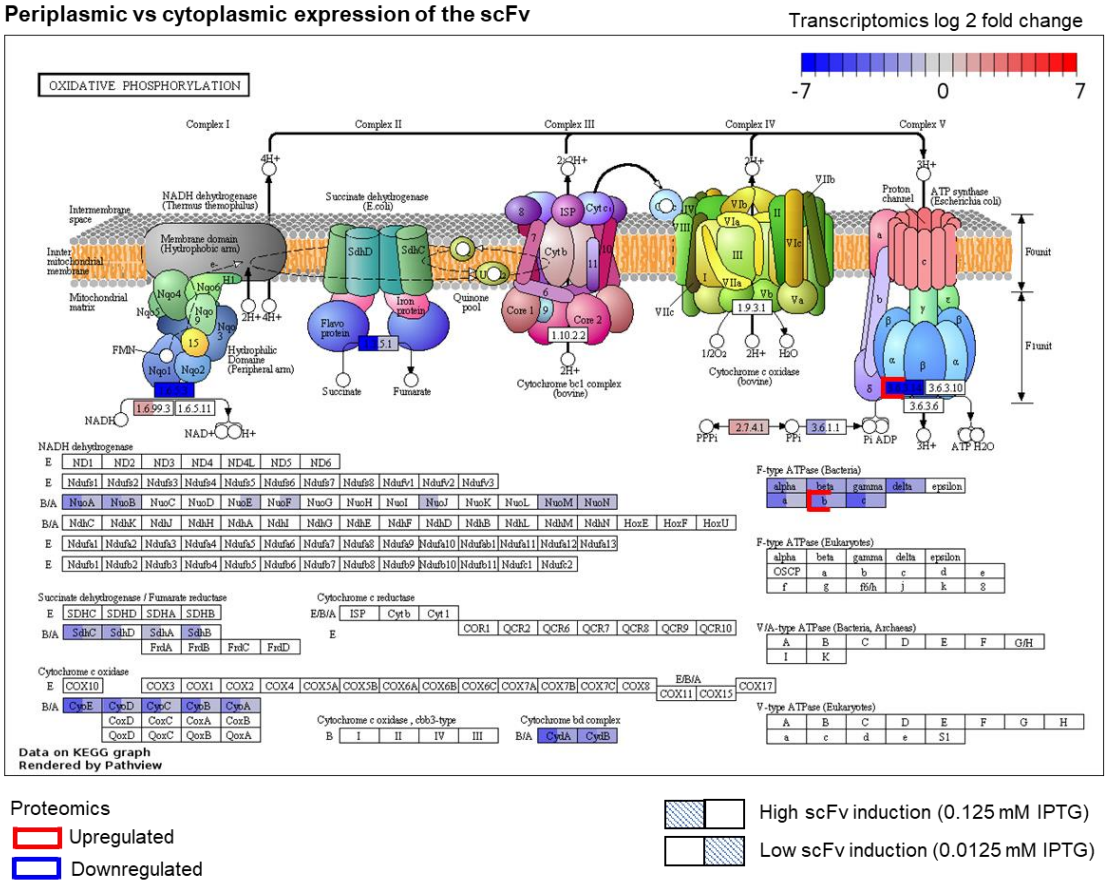


Figure 62. KEGG map of oxidative phosphorylation pathway (ecj00190) containing differentially expressed genes and proteins when comparing expression of the scFv in the periplasm and cytoplasm. Samples were generated when inducing either high (left) or low (right) levels of scFv expression. Differentially expressed proteins are circled in red if they are upregulated and blue if they are downregulated. Transcriptomic log2 fold changes of differentially expressed transcripts are coloured in red if upregulated and blue if downregulated. Data was visualised using Pathview [275], [276].

3.3.3.5.3.2 Fatty acid synthesis and degradation

ScFv secretion led to an increase in expression of genes involved in fatty acid degradation and a decrease in expression of genes involved in fatty acid

synthesis at both the proteomic and transcriptomic scale, regardless of the amount of inducer added (Figure 63, Appendix 11). Fatty acids play many important roles in the cell including in cell wall formation and biotin formation (which was also identified as downregulated upon high induction based on KEGG enrichment (Figure 51)).

Indeed this is consistent with other transcriptomics studies that found that overexpressing a different antibody fragment in the periplasm led to a decrease in expression of genes involved in fatty acid metabolism indicating that this response may not be unique to the scFv analysed here [273].

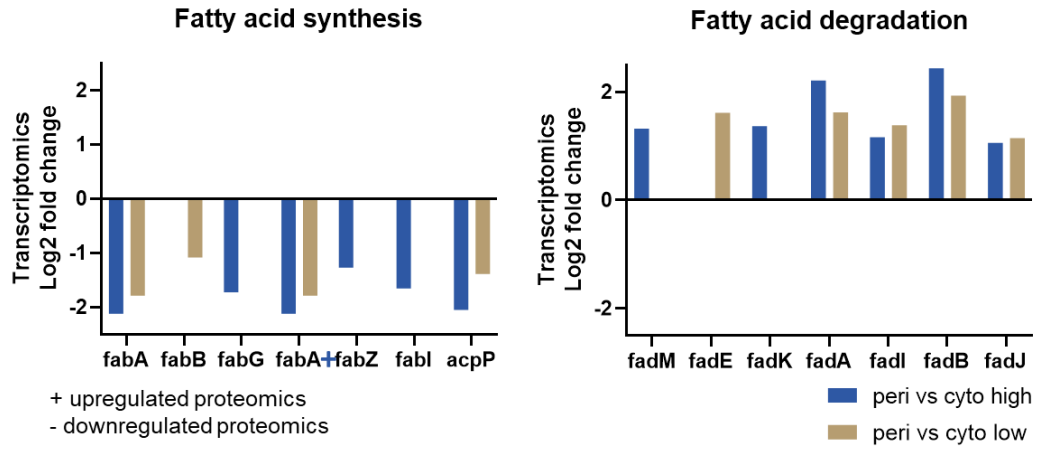


Figure 63. Differentially expressed genes and proteins involved in fatty acid synthesis and degradation when comparing cell expressing the scFv in the periplasm (peri) and cytoplasm (cyto) when induced with high (0.125 mM IPTG) and low (0.0125 mM yellow) levels of IPTG. Transcriptomic log2 fold changes are indicated as well as upregulated proteins (plus sign) and downregulated proteins (minus sign). Genes and proteins were assigned to fatty acid synthesis and degradation pathways based on classifications in the Ecocyc OMICs dashboard [252].

3.3.3.5.3.3 Amino acid synthesis and degradation

Secretion of the scFv to the periplasm affected expression of genes involved in amino acid synthesis and degradation as identified through enrichment analysis (Figure 51, Appendix 12, Appendix 14, Appendix 15, Appendix 21, Appendix 22, Appendix 24). Amino acids are the building blocks of proteins and therefore play an important role in cell physiology and recombinant protein overexpression.

There was an increase in expression of alanine, isoleucine, leucine and phenylalanine synthesis genes when comparing cells expressing the scFv in the periplasm (with OmpA signal peptide) and cytoplasm (no OmpA signal peptide) (Figure 64). These amino acids are in high abundance in the OmpA signal peptide (8, 2, 1 and 1 of each amino acid respectively) which could potentially explain this increase.

There was also an increase in expression in amino acid synthesis pathways involved in sulphur metabolism upon scFv secretion including serine, glycine, threonine and cysteine synthesis pathways (Figure 64). Indeed, sulphur metabolism was identified as enriched amongst transcriptomics data and the CysB regulon was upregulated upon high levels of scFv secretion based on TFinfer analysis (Figure 50). This indicates that strains may be lacking sulphur and is consistent with the increase in expression of genes involved in iron sulphur cluster assembly seen in section 3.3.3.5.1.3.

ScFv secretion led to a decrease in the synthesis and an increase in the degradation of certain amino acid pathways. Proline and methionine synthesis pathways were downregulated at the transcript level and arginine and histidine pathways at a protein level (Figure 50). Indeed, there was an increase in the MetJ regulon (a negative regulator of methionine synthesis) based on TFinfer analysis. However, glycine, proline and serine degradation pathways were upregulated at the transcriptomic scale, particularly when protein production was induced with high levels of IPTG that overloaded the secretion (Figure 50). These amino acids make up a large percentage of the scFv (around 8-15% of the scFv each) and therefore a decrease in their expression could have a negative impact on translation of the scFv. Further metabolomics would, however, need to be conducted to confirm this and determine the abundance of these amino acids in the cell.

Multiple other OMICs studies have also found that recombinant protein overexpression affects expression of amino acid synthesis pathways and that this response varied based on the amino acid composition of the recombinant protein that was overexpressed [201], [203], [211], [216], [232], [285]–[287].

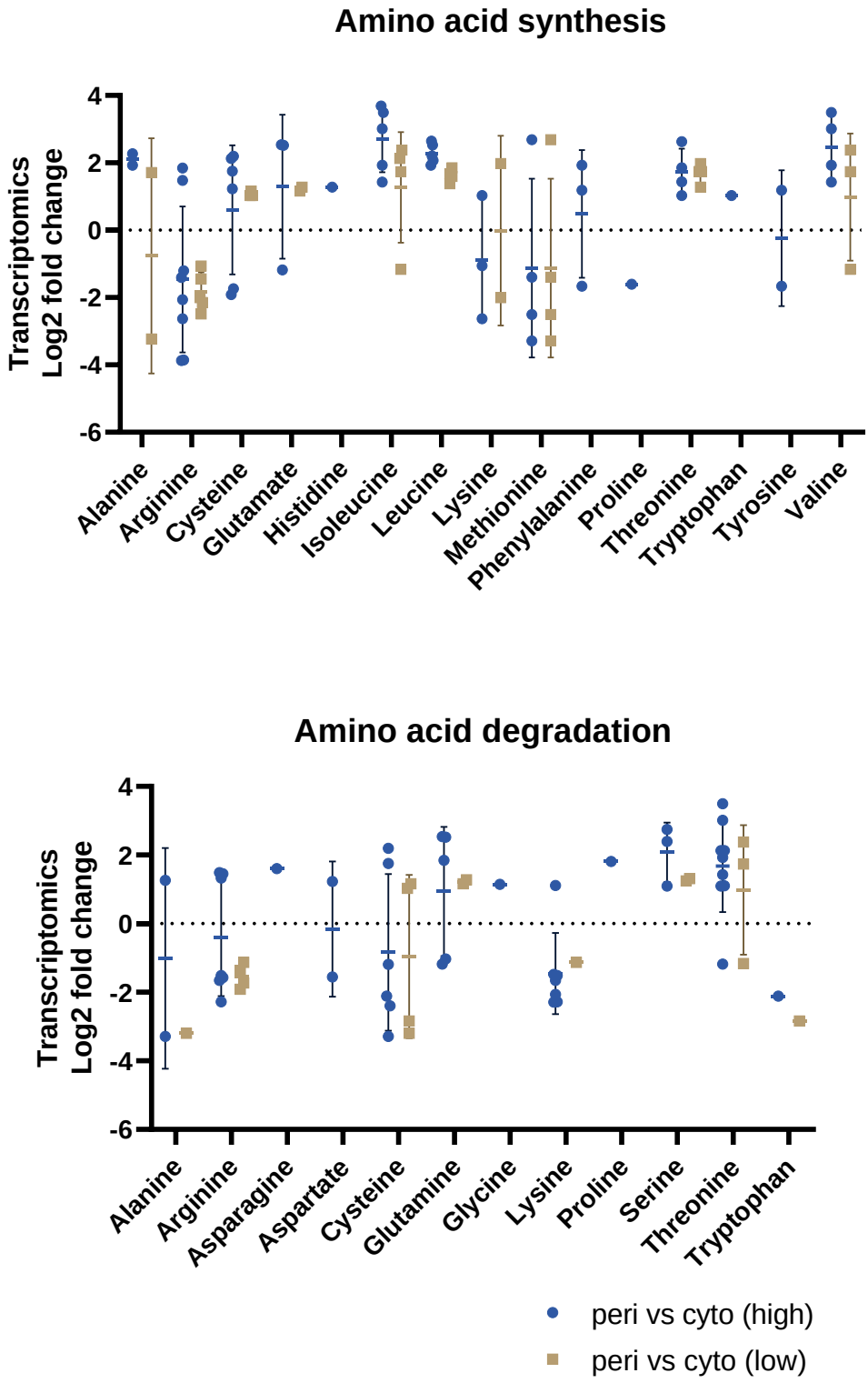


Figure 64. Differentially expressed transcripts involved in amino acid synthesis and degradation when comparing strains expressing an scFv in the periplasm (peri) or cytoplasm (cyto) when induced with high (0.125 mM blue) or low (0.0125 mM yellow) concentrations of IPTG. Each dot represents a differentially expressed transcript and the mean fold change and standard deviation (error bars) are also represented. Genes were grouped based on Ecocyc OMICs dashboard classifications [252].

3.3.3.5.3.4 Ribosome synthesis

ScFv secretion led to a decrease in expression of genes involved in ribosome synthesis at a transcriptomic scale particularly when high levels of scFv were secreted which overloaded the Sec secretion system (Figure 65, Figure 48, Figure 51, Appendix 17, Appendix 18, Appendix 21, Appendix 22). This involved a decrease in both small and large ribosomal subunit proteins which are responsible for ribosome assembly as well as the elongation factors EF-Tu which enable binding of aa-tRNA to the ribosome and protein synthesis were affected by scFv secretion. Interestingly very few ribosomal subunits were differentially expressed at the protein level.

This is consistent with previous OMICs studies that also found that recombinant protein overexpression, secretion and secretion system overloading led to a decrease in expression of ribosomal proteins [90], [212], [215], [216] however some studies have found opposing results [273]. This indicates that differences in expression of ribosome synthesis genes may be dependent on culture conditions, properties of the overexpressed protein and/or the level of secretion system overloading.

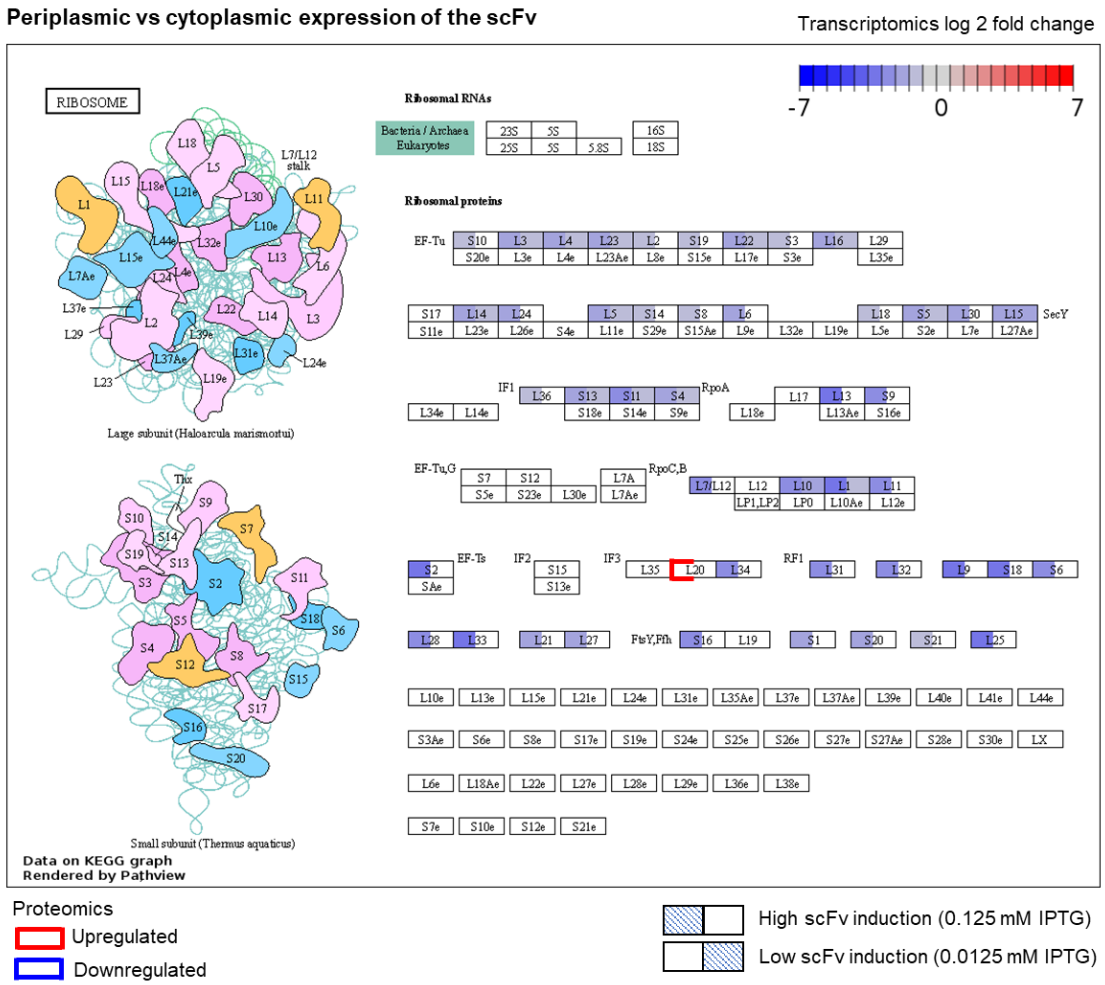


Figure 65. KEGG map of ribosome subunits (ecj03010) containing differentially expressed genes and proteins when comparing expression of the scFv in the periplasm and cytoplasm. Samples were generated when inducing either high (left) or low (right) levels of scFv expression. Differentially expressed proteins are circled in red if they are upregulated and blue if they are downregulated. Transcriptomic log2 fold changes of differentially expressed transcripts are coloured in red if upregulated and blue if downregulated. Data was visualised using Pathview [275], [276]**.**

3.4 Conclusion – Cellular responses to recombinant protein secretion

Strains were successfully cloned to express the scFv in the cytoplasm and periplasm (adding OmpA signal peptide) under industrially relevant high cell density cultures. Transcriptomics (RNA-seq) and proteomics (iTRAQ) of these samples allowed us to gain a better understanding of cellular responses to scFv secretion, the effect of secretion system overloading (by comparing strains

induced with different concentrations of IPTG) and how these cellular responses changed over time (by comparing strains 12- and 24-hours post-induction).

Secretion of the scFv had a significant impact on the transcriptome and proteome, with high levels of scFv expression that overloaded the secretion system having a more severe effect on the cell than lower levels of scFv expression that overloaded the Sec secretion system less.

ScFv secretion led to an increase in several stress responses including the heat shock response, iron stress, oxidative stress and the SOS response. Secretion of the model recombinant protein also led to an increase in certain envelope stress response pathways including the PspA, RpoE and CpxR regulons however there was only an increase in the CpxR response was only seen when inducing low levels of scFv secretion that overloaded the secretion system less.

The envelope stress response regulates expression of several membrane and periplasmic protein folding genes whereas the heat shock response predominantly regulates expression of cytoplasmic protein folding genes. Indeed, there was also an increase in expression of several cytoplasmic periplasmic and membrane chaperones and proteases upon scFv secretion. Interestingly however there was a decrease in expression of several periplasmic protein folding genes upon scFv secretion which could be detrimental to scFv folding. ScFv secretion also negatively affected expression of several transcripts that play an important role in Sec Secretion and negatively impacted expression of envelope proteins when the secretion system was overloaded. ScFv secretion also led to an increase in expression of pili and flagella genes.

ScFv secretion also had an effect on metabolism leading to a decrease in expression of several genes involved in oxidative phosphorylation, a decrease in fatty acid synthesis, an increase in fatty acid degradation and a decrease in ribosome synthesis. ScFv expression also affected expression of multiple genes involved in amino acid synthesis and degradation. Secretion of the model protein led to an increase in glycine, proline and serine degradation pathways which may affect scFv expression as these amino acids make up a large percentage of the scFv (each around 8-15% of the scFv).

This OMICs data was then used to guide genetic engineering of strains that will be tested to determine if they improved secretion of the scFv to the periplasm under shake flask and industrially relevant conditions as well as production of other periplasmic recombinant proteins which will be covered in the following chapter.

4 Chapter 4. Engineering strains and media to improve recombinant protein secretion to the periplasm

4.1 Intro

Multiple studies have used genetic engineering to improve production and secretion of recombinant proteins to the periplasm by altering expression of periplasmic and cytoplasmic protein folding genes, increasing expression of the Sec or TAT secretion system, and increasing expression of stress response genes (see introduction section 1.1.4.3).

However, this approach often only improves production of specific recombinant proteins [99] and relies heavily on prior knowledge about cellular responses to recombinant protein overexpression and secretion which is limited by the fact that 1/3rd of *E. coli* genes have no known function [288].

Therefore, OMICs analysis has not only been a useful tool for understanding cellular responses but has also been used to guide strain and media alterations (see introduction section 1.2.3.4). Using OMICs presents several advantages in that it is unbiased and does not rely on prior knowledge about gene function and can be integrated with metabolic modelling to identify key genes and bottlenecks. OMICs data generated in Chapter 3 was therefore used to direct strain and media engineering.

This chapter will start by covering the development of a scale down model in shake flasks to test engineered strains and media. It will then move on to describe the rationale behind alterations to strains and media made in this study and methods used to genetically engineer strains. The final section will include testing strains and media for improved production and secretion of the scFv and other model therapeutic proteins (Hel4, sfGFP and HGH) under shake flask and industrially relevant HCDC conditions.

4.2 Aims

- Establish a scale-down model in shake flasks to replicate Ambr fermenter conditions in a high throughput manner.
- Engineer strains and media based on knowledge from OMICs data described in Chapter 3.
- Test modified strains and media for improved production and secretion of the scFv and other model therapeutic proteins (Hel4, sfGFP and HGH) under shake flask and industrially relevant HCDC conditions.

4.3 Results and discussion

4.3.1 Establishing scaled down models in shake flasks

4.3.1.1 Growing cultures in shake flasks.

Ideally, strain and media alterations would have been tested under the same conditions used to generate OMICs data (Ambr bioreactors/high cell density cultures) as changing culture conditions is known to affect cellular responses [169], [170], [204], [223], [232], [234]. However, growing cultures in bioreactors in a high throughput manner is time consuming, expensive and requires specialised equipment. Therefore, in this study, a scale down model was developed in shake flasks to test alterations to strains and media.

Baffled shake flasks were chosen as they had increased aeration which is known to be an important factor when achieving high cell densities [289]. Bacteria were grown in media similar media to that used in Ambr bioreactors (M9 salts, 20 g/L yeast extract and 35 g/L glycerol). Media was similar but not exactly the same as that used in Ambr fermenters as we were not aware of specific Ambr conditions at the time. 50 mL cultures were grown in 250 mL flasks and were incubated 37 °C with shaking (150 rpm). Growth of the wildtype W3110 Δ*ompT* was initially tested to determine the maximum cell density achieved under these conditions and monitor how media pH changed over time. Under these conditions exponential growth occurred during the first 4 hours of incubation with doubling times taking on average 35 mins and with the pH remaining constant at pH 7. After 4 hours, there was a drop in pH as cultures

entered stationary phase which reached pH 4 after growth overnight. The maximum cell density achieved was around an OD_{600} of 9 (Figure 66). This maximal cell density is a lot higher than densities achieved using other media such as LB (maximum OD_{600} of around 7), which is known to be limited due to a lack of carbon source [290]

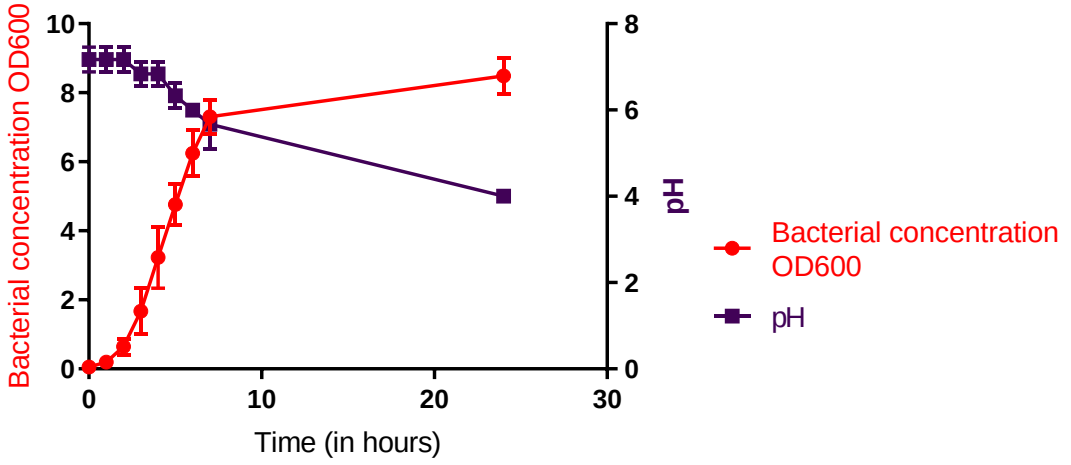


Figure 66. Growth curve of E. coli monitoring pH. Error bars represent 3 technical replicates. Results are representative of 3 technical repeats. Bacterial concentrations were measured based on absorbance and pH of the supernatant was measured using a colorimetric indicator. Error bars represent standard deviation.

4.3.1.2 Improving growth under shake flask conditions

To further increase bacterial cell densities to replicate high cell density culture conditions, multiple hypotheses were formed about factors that were limiting growth. These were tested by adding carbon sources to media (2% w/w glycerol), adding nitrogen sources to media (6-6.6% ammonium hydroxide), altering media pH (adding 6-6.6% ammonium hydroxide or 100mM solution of sodium hydroxide to adjust the pH to 7, 1x MOPS buffer) and improving aeration (reducing culture volumes to 20 mL). Glycerol, nitrogen, ammonium hydroxide and sodium hydroxide were all added 5 hours into fermentation where the pH dropped, and growth started to plateau.

Adding carbon and nitrogen did not affect growth (Figure 67 A and B). Adding ammonium hydroxide, sodium hydroxide or MOPS, reduced acidification of media over time, but also had no effect on growth. However, addition of MOPS buffer had a negative impact on growth within the first 5 hours of growth (Figure

67, B, C and D) (2-way ANOVA n=3). Reducing culture volumes to 20 mL to improve aeration did however lead to an increase in the final bacterial concentration of *E. coli* with the bacterial concentrations significantly increasing 6 hours into growth (2-way ANOVA) to reach an OD_{600} of around 13 (Figure 67 E). This indicates that aeration may be one of the main limiting factors in shake flask cultures and therefore culture volumes in shake flasks will be reduced to 20 mL in all future work. Other studies have also found that reducing culture volume improved overall bacterial densities [289].

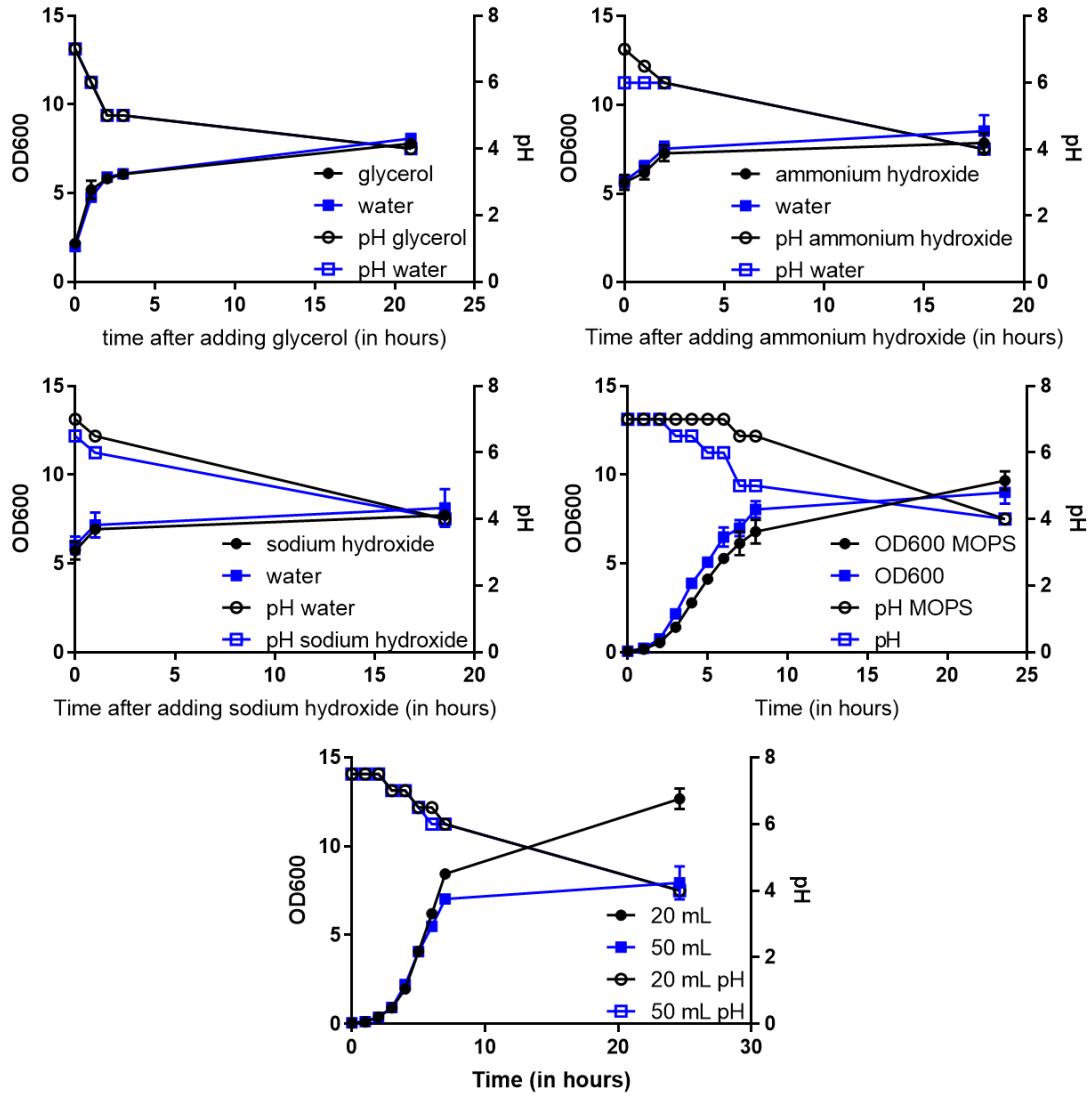


Figure 67. Growth curves monitoring the effect of adding glycerol (A), adding nitrogen (B), adjusting the pH of media (B, C and D) and improving aeration (E) on growth and pH. Error bars represent 3 technical replicates. Error bars represent standard deviation. Significantly different growth was determined using a 2-WAY ANOVA and is represented by a star.

4.3.1.3 Overexpressing the scFv under shake flask conditions

Once optimal growth conditions were established, expression of strains was induced in shake flasks containing 20 mL of media that was more similar to that used in Ambr fermentions. Induction of scFv expression was conducted mid log phase (at an OD_{600} 3-4). Expression of the scFv in the cytoplasm had no significant effect on growth compared to the wildtype (Figure 68). However, expression of the scFv in the periplasm lead to a decrease in growth once induced (Figure 68). This is probably due to cell lysis or leakiness due to high levels of scFv accumulation in the periplasm as can be seen by the increased quantity of whole cell proteins in these samples 3 hours post induction based on SDS page (Figure 69). This differs from high cell density cultures where bacteria started to lyse 24 hours post induction (see section 3.3.2.2). SDS-PAGE analysis of whole cells can also be found in Appendix 7.

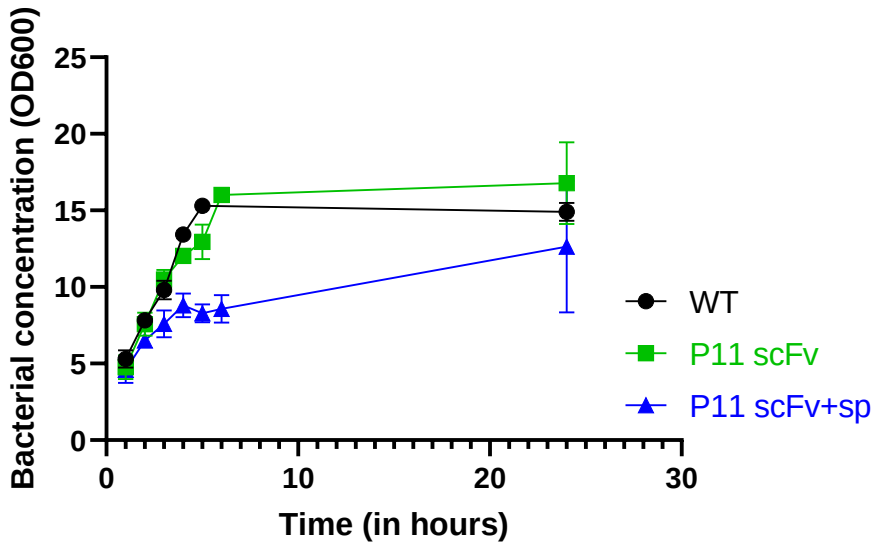


Figure 68. Effect of overexpressing an scFv in the cytoplasm (P11 scFv) and the periplasm (P11 scFv+sp) on growth under shake flask conditions compared to an empty vector (P11) control. Samples were induced with 0.125 mM IPTG at an OD_{600} 4-6. N=3, error bars represent 1 SD.

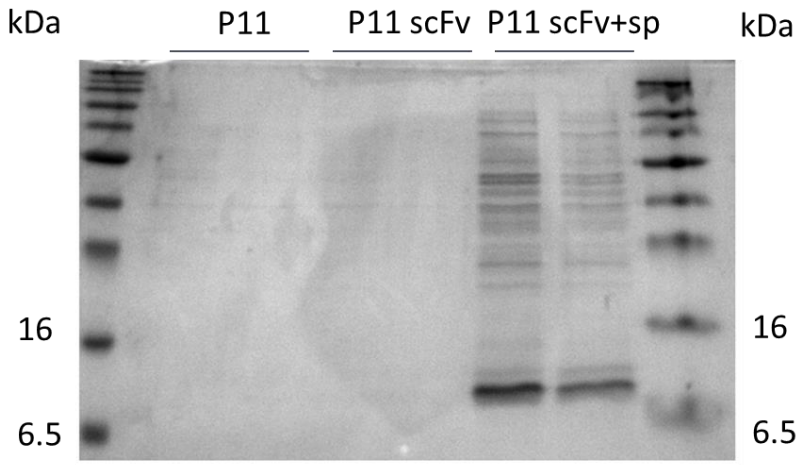


Figure 69. SDS-PAGE of supernatant samples when expressing the scFv in the cytoplasm (P11 scFv) and periplasm (P11 scFv+sp), 3 hours post induction with 0.125 mM IPTG under shake flask conditions. P11 is an empty vector induced control. The scFv has a molecular weight of 14 kDa with and 12 kDa without the signal peptide. N=2.

4.3.2 Designing alterations to media and strains.

Based on OMICs data, six key areas that were affected by scFv secretion were identified for targeted modifications to improve protein secretion which are summarised in Figure 70. Multiple considerations were taken when deciding what genes and pathways were altered. This included whether the gene was identified as differentially expressed at both the protein and transcript level, whether a gene was identified as differentially expressed amongst the multiple conditions tested, whether the protein was identified using mass spectrometry, whether multiple genes in said pathway were identified as differentially expressed, their level of expression and information from the literature.

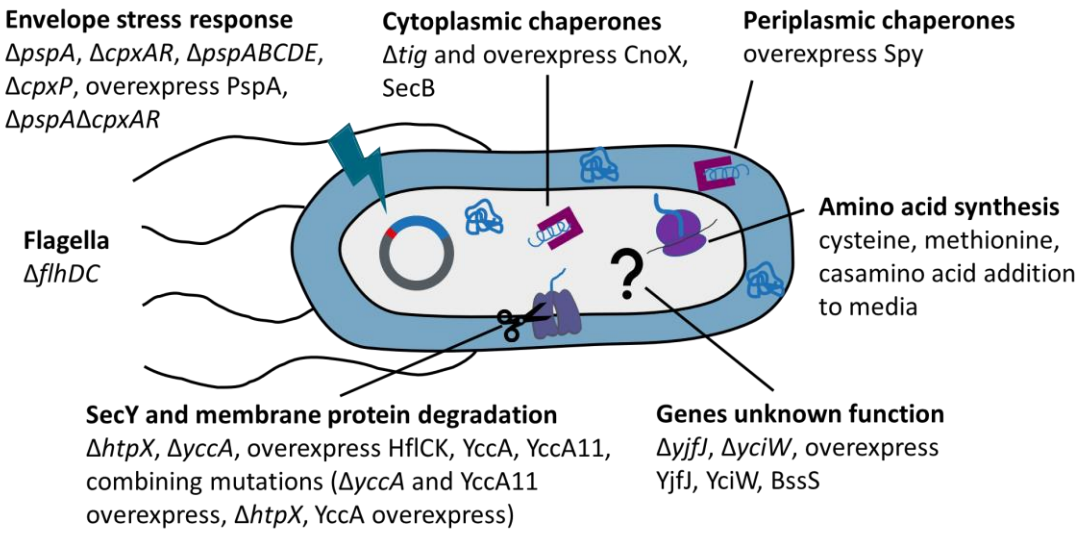


Figure 70. Figure summarising all engineered strains and media used as part of this study.

4.3.2.1 Envelope stress response

ScFv secretion affected expression of genes involved in the envelope stress response (see section 4.3.2.1). The envelope stress response responds to unfolded periplasmic and membrane proteins and is activated by multiple pathways (CpxAR, BaeSR, Ras, RpoE and PspA) that regulate expression of genes involved in periplasmic and membrane protein folding [65], [66], [78]–[81]. A hypothesis was therefore formed that overexpressing envelope stress response pathways would improve production and secretion of recombinant proteins (including the scFv) by alleviating stress and improving recombinant protein folding in the periplasm.

ScFv secretion led to an increase in the *psp* response and PspA expression, particularly when high levels of scFv were secreted (0.0125 mM IPTG induction), which was more pronounced over time, however, no changes in PspA expression were seen at the protein level. PspA is both a negative regulator of the *psp* operon and binds to damaged inner membranes to help maintain proton motive force which plays a critical role in secretion (Figure 71) [87], [88], [291], [292]. Furthermore, multiple studies have previously identified that overexpressing PspA improved yields of secreted recombinant proteins. The effect of overexpressing PspA on expression and secretion of our scFv was therefore

tested. *PspA* and the *psp* operon (*pspABCDE*) were also knocked out to determine if this was detrimental to recombinant protein secretion.

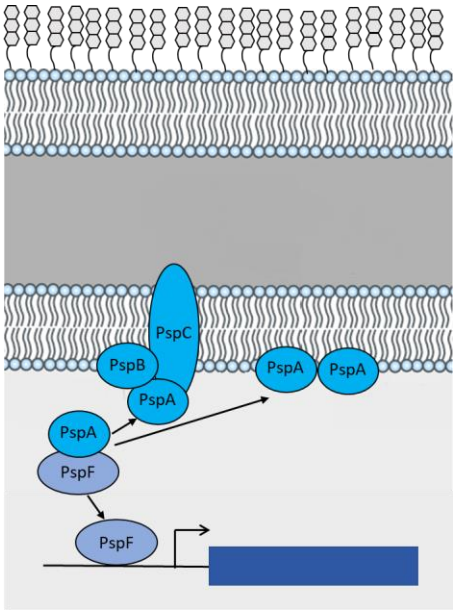


Figure 71. Figure representing regulation and activity of the psp response. PspF regulates expression of the psp operon and is negatively regulated by PspA when it is not bound to PspB and PspC. PspA also forms an oligomeric ring and protects cells from membrane damage by binding damaged membranes and maintaining proton motive force. Figure adapted with permission from [82]***.***

ScFv secretion also led to a decrease in the *cpx* response, particularly when inducing low levels of scFv expression (0.0125 mM IPTG induction). The *cpx* response is known to respond to unfolded membrane proteins and regulate expression of genes involved in periplasmic protein folding and membrane protein degradation (Figure 72) [78], [84], [269], [270], [293]. The *cpx* response is activated post translationally by the CpxAR two-component system and is activated by phosphorylation of CpxR by CpxA (Figure 72). Therefore, increasing expression of CpxR will not necessarily increase the *cpx* response. In order to increase activity of the *cpx* response, the negative regulator *cpxP* was knocked out and tested for improved scFv expression and secretion (Figure 72) [269], [294]. CpxAR was also knocked out to see if removing this pathway had a detrimental effect on recombinant protein production and secretion. Strains were also generated that combined these mutations to inhibit both pathways (*ΔpspA* and *ΔcpxAR*).

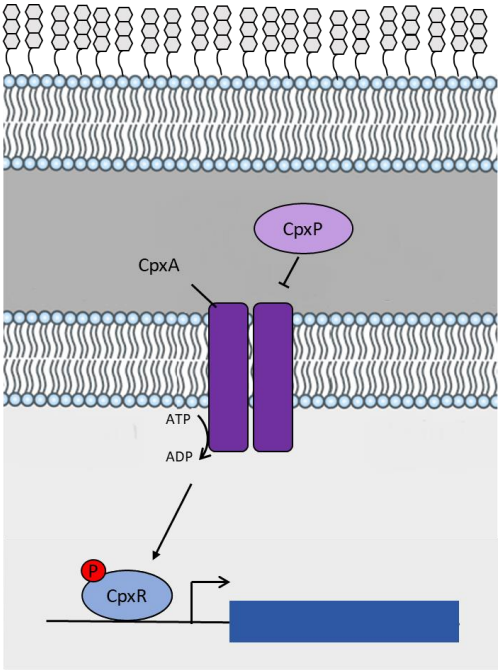


Figure 72. Figure representing activity and regulation of the cpx response. CpxA phosphorylates and activates CpxR which binds to DNA regulating transcription. CpxP is a negative regulator of the Cpx response. Figure adapted with permission for [82]

4.3.2.2 Periplasmic chaperones

ScFv secretion affected expression of multiple genes involved in periplasmic protein folding (see section 3.3.3.5.2.2). There was a decrease in expression of the periplasmic chaperone *spy* at the transcriptomic level when high levels of scFv expression in the periplasm were induced (0.125 mM IPTG). No difference in spy expression was however seen at the proteomics level. Spy plays an important role in periplasmic protein folding. Therefore, a hypothesis was formed that increasing spy expression would improve the folding and yields of recombinant periplasmic proteins such as the scFv.

4.3.2.3 Cytoplasmic chaperones

Secreting the scFv also affected expression of multiple genes involved in cytoplasmic protein folding (see section 3.3.3.5.2.2). These chaperones can be beneficial or detrimental to protein secretion. The effect of knocking out and overexpressing several cytoplasmic chaperones was therefore tested.

CnoX was shown to be upregulated at both the proteomics and transcriptomics level upon scFv secretion particularly when the secretion system

was overloaded (high induction 0.125 mM IPTG). CnoX is a chaperedoxin and protects cysteines from oxidation in the cytoplasm [295] and therefore may play an important role in preventing the folding of recombinant periplasmic proteins (which often contain cysteines to form disulphide bonds) in the cytoplasm. Therefore, a strain overexpressing CnoX was tested as part of this study for improved scFv production and secretion.

High levels of scFv secretion (0.125 mM IPTG) also led to a decrease in expression of *secB* transcription. SecB is a chaperone that plays an important role in secretion binding to preproteins, guiding them to SecA and the SecYEG secretion system [58]. SecB was therefore overexpressed to determine if this further improved production and secretion of the scFv.

ScFv secretion led to an increase in expression of the secretory chaperone Tig at both the RNA and protein level when induced with both high (0.125 mM) and low (0.0125 mM) concentrations of IPTG. Tig binds to secreted preproteins preventing their folding in the cytoplasm. However, Tig has also been shown to delay protein secretion [130], [131]. Therefore, the effect of knocking out Tig was tested to see if this improved scFv production and secretion.

4.3.2.4 SecY and membrane protein degradation

ScFv secretion led to an increase in expression of the two membrane proteases FtsH transcription and HtpX (at a proteomics and transcriptomics level) (see section 3.3.3.5.2.2). These membrane proteases are known to degrade SecY a core component of the Sec translocon, through which the recombinant scFv was secreted (Figure 73) [279], [296]. Indeed, jamming of the secretion system has been shown to increase degradation of SecY and other transporters, having a detrimental effect on the cell [280]. A hypothesised was therefore formed that inhibiting membrane protein degradation would increase SecYEG levels, increasing secretion of the scFv and relieving stress associated with high levels of membrane protein degradation.

The membrane protease HtpX was therefore knocked out and tested for improved scFv secretion. It is not possible to knockout the membrane protease FtsH as it is an essential gene [297], [298]. However overexpressing negative regulators of FtsH (HflCK, YccA and YccA11 (a truncated form of YccA which is

more stable and active)) has been shown to reduce FstH activity and reduce SecY and membrane protein degradation (Figure 73) [280], [281], [299], [300]. HflCK was also upregulated and YccA downregulated at the transcript level upon scFv secretion. The effect of overexpressing HflCK, YccA and YccA11 on scFv production and secretion was therefore tested. The effect of combining mutants to further inhibit membrane protein degradation (Δ*htpX*, YccA overexpression) was also tested.

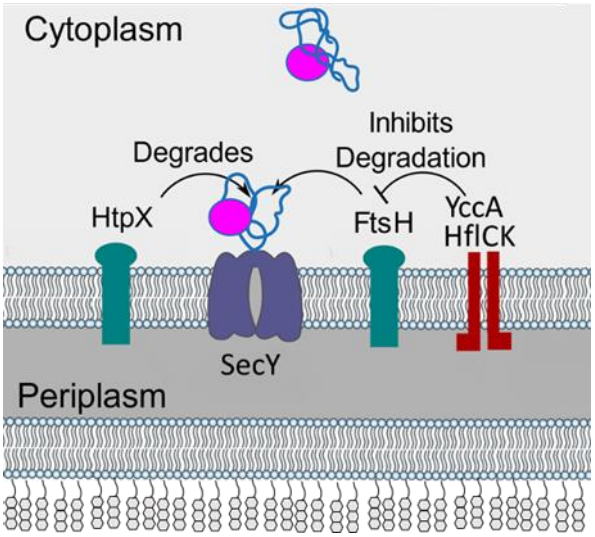


Figure 73. Figure representing proteins involved in degrading SecY and other membrane proteins (HtpX and FtsH) as well as proteins that regulate their activity (YccA and HflCK).

4.3.2.5 Amino acid addition

ScFv secretion had a negative impact on expression of multiple genes involved in amino acid synthesis including an increase in the CysB regulon (positive regulator of cysteine biosynthesis) and the MetJ regulon (negative regulator of methionine biosynthesis) (see section 3.3.3.5.3.3). To determine if amino acid limitations were affecting scFv production and secretion, cysteine, methionine or casamino acids (mixture of amino acids) were added to media and scFv expression and secretion were tested.

4.3.2.6 Flagella

ScFv secretion also led to an increase in expression of several genes and proteins involved in flagella synthesis, particularly when the secretion system was overloaded due to high levels of scFv expression (0.125 mM IPTG) (see

section 3.3.3.5.2.4). This increase in flagella synthesis could be detrimental to scFv secretion as it uses up cellular resources and relies on proton motive force which is important for Sec secretion. Furthermore, the strains used in this study were non motile and flagella synthesis was not necessary (Appendix 26). The effect of knocking out the master regulator of flagella synthesis FlhDC was therefore tested to see if it improved secretion of the scFv.

4.3.2.7 Genes of unknown function

Multiple genes of unknown function were identified as being differentially expressed upon scFv secretion. Indeed around 1/3rd of *E. coli* genes have no known function [288]. The effect of knocking out and overexpressing these genes was therefore tested to see if this improved scFv section and further determine their function.

YceP (BssS) was identified as being upregulated upon scFv secretion regardless of the amount of IPTG added (0.125 or 0.0125 mM IPTG) at the proteomics level. BssS is known to regulate expression of around 400 genes based on transcriptomics analysis [301]. BssS is thought to play a role in biofilm formation however there are contradictory reports as to whether it is a positive or negative regulator of biofilm formation [301]–[304]. The effect of overexpressing BssS on the production and secretion of the scFv, was therefore tested.

There was an increase in *yjfJ* transcription upon scFv secretion, regardless of the amount of IPTG added (0.125 or 0.0125 mM IPTG). *YjfJ* is a homologue of the envelope stress response gene *pspA* [305] which has previously been shown to play an important role in maintaining proton motive force and protein secretion [87], [88], [146]–[149], [291], [292]. It was therefore hypothesised that YjfJ also played an important role in secretion and therefore the effect of knocking out and overexpressing YjfJ on scFv expression was also tested.

YciW was also another gene of unknown function that was upregulated upon scFv secretion regardless of the amount of IPTG added (0.125 or 0.0125 mM IPTG) at a proteomics and transcriptomics scale. *YciW* has no known function but is thought to play a role in cysteine synthesis which was also identified as changing upon scFv secretion (see section 3.3.3.5.3.3) [306], [307].

The effect of knocking out and overexpressing YciW on scFv expression and secretion was therefore tested.

4.3.3 Constructing and verifying mutants and alteration to media

Based on OMICs data, 10 genes were overexpressed, 10 genes were knocked out, 3 strains with combined mutations and 3 alterations to media were generated to test if they improved scFv secretions. The following section will cover methods used to overexpress genes, knockout genes and alter media as well as validation and testing of modifications to see if they had an impact on growth.

4.3.3.1 Overexpressing genes

4.3.3.1.1 Cloning constructs

Multiple methods were trialled to coexpress genes at different levels. Genome integration is considered the gold standard as it requires less antibiotic usage and therefore a pDOC system that uses lambda red recombineering was trialled to integrate constitutive Anderson promoters of different strengths into the genome [308]. However, this process was timely and therefore plasmids were used instead to overexpress proteins. An arabinose inducible plasmid (pBAD33) was trialled in a Δ*araBA* background strain provided by FDB (as IPTG was already being used to induce scFv expression). However, this presented several disadvantages since the presence of this plasmid had a negative impact on strain growth (potentially due to high levels of expression of the chloramphenicol resistance gene), had limited tunability and expressed too high quantities of our protein of interest (Appendix 27).

Proteins were therefore constitutively overexpressed by cloning our gene of interest into pAS15a-cm plasmid containing Anderson promoters of different strengths (iGEM part BBa_J23119 (high), BBa_J23106 (medium) or BBa_J23103 (low)) [238]. pAS15a-cm was specifically designed for this study, was synthesised by Genewiz and digested/ ligated to form a plasmid using BamHI. The plasmid incorporated several features to be compatible with strains and plasmids used in this study (Figure 74):

1- A p15a origin of replication that is compatible with the ColE1/pMB1/pBR322/pUC origin of replication in the Paveway 11 plasmid (used to overexpress the scFv) which has a low copy number (around 10 copies per cell).

2- A chloramphenicol antibiotic resistance selection marker (CmR) that is compatible with all strains and the Paveway 11 plasmid.

4- A Multiple Cloning Site (MCS) with multiple restriction sites to clone constitutive Andersson promoters and our gene of interest.

3- Terminators at the 3' sites of the MCS and the CmR to stop transcription.

Cloning of Anderson promoters and genes of interest was conducted using two steps (Figure 74). First, primers containing one of the three Anderson promoters and a ribosome binding site (BBa_J61100) were annealed to make double stranded DNA (see example in Figure 75) which was then cloned into pAS15a-cm using SpeI and EcoRI restriction sites and verified using colony PCR and sequencing (see example in Figure 76). In the second step, the gene of interest (in this example *pspA*) was amplified from genomic DNA using PCR, cloned into the vector using SpeI, PstI or XmaI restriction sites and verified using colony PCR and sequencing (Figure 77).

The *pspA* gene was cloned into plasmids containing all three Anderson promoters (high, medium and low) as this was the first gene studied. Due to time restrictions, all other genes of interest were synthesised and cloned into vectors by Genewiz and proteins were only overexpressed with two promoters to reduce costs (see sequences in Appendix 1). All overexpression plasmids used high and medium strength promoters other than those overexpressing Spy that used a medium and low promoter so as to not to further overwhelm the Sec secretion system as Spy is secreted via this pathway.

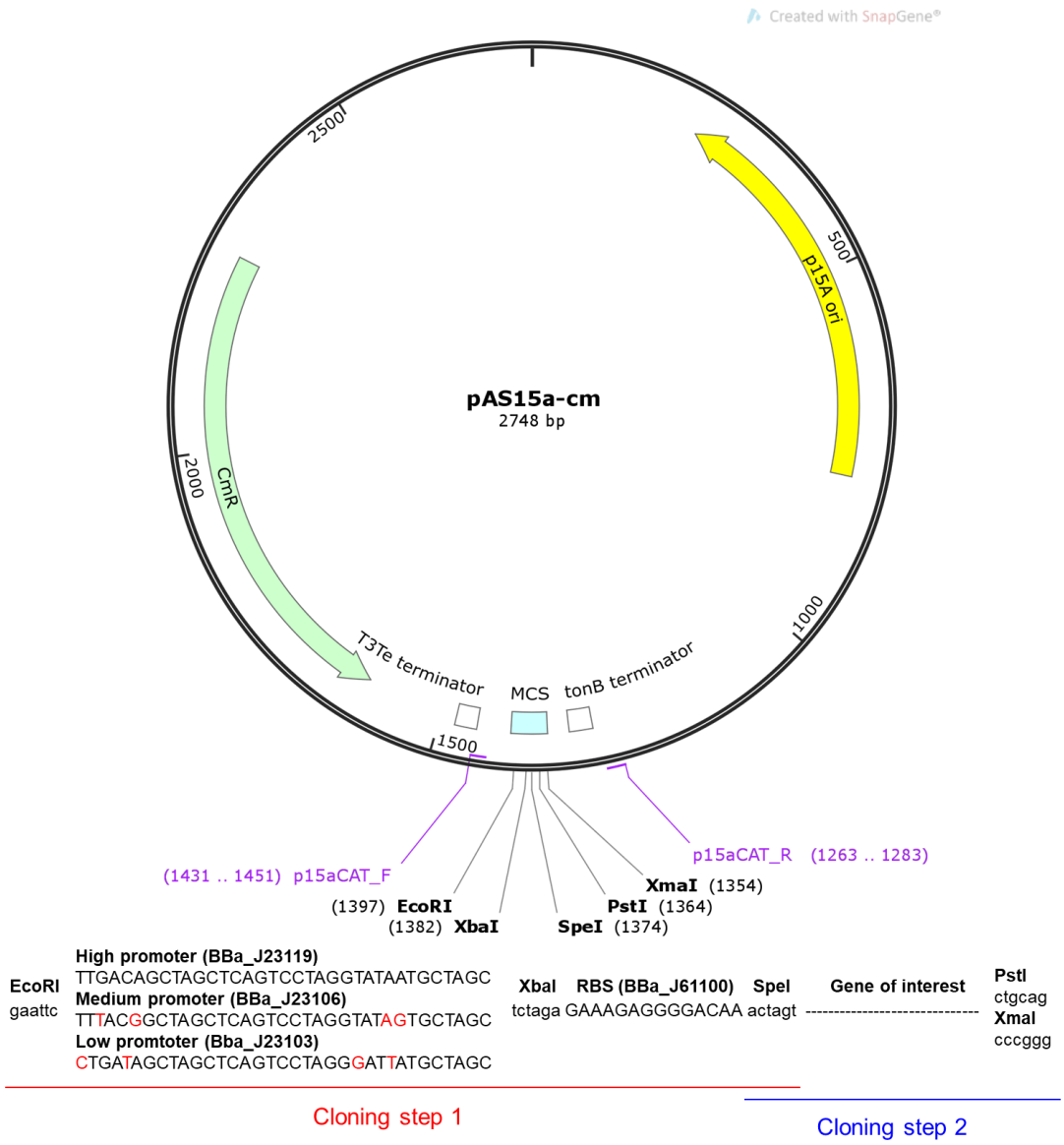


Figure 74. Figure representing cloning of high, medium and low Anderson promoters and our gene of interest into pAS15a-cm plasmid. The plasmid map was created using SnapGene with ori representing the origin of replication, CmR the chloramphenicol resistance cassette, MCS the multiple cloning site and p15a_CAT_F and p15a_CAT_R are primer binding sites were used to check cloning. Cloning was conducted in two steps: One that integrated an Anderson promoter and ribosome binding site into the plasmid and a second step that integrated the gene of interest. Differences between promoter sequences are indicated in red.

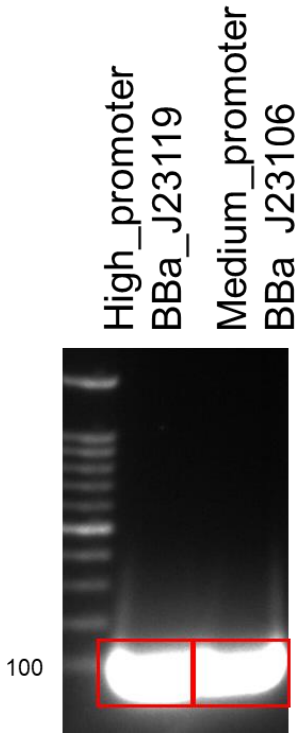


Figure 75. 2% gel containing annealed oligos used to make double stranded DNA containing Anderson promoters. Annealed oligos are 70 bp and circled in red.

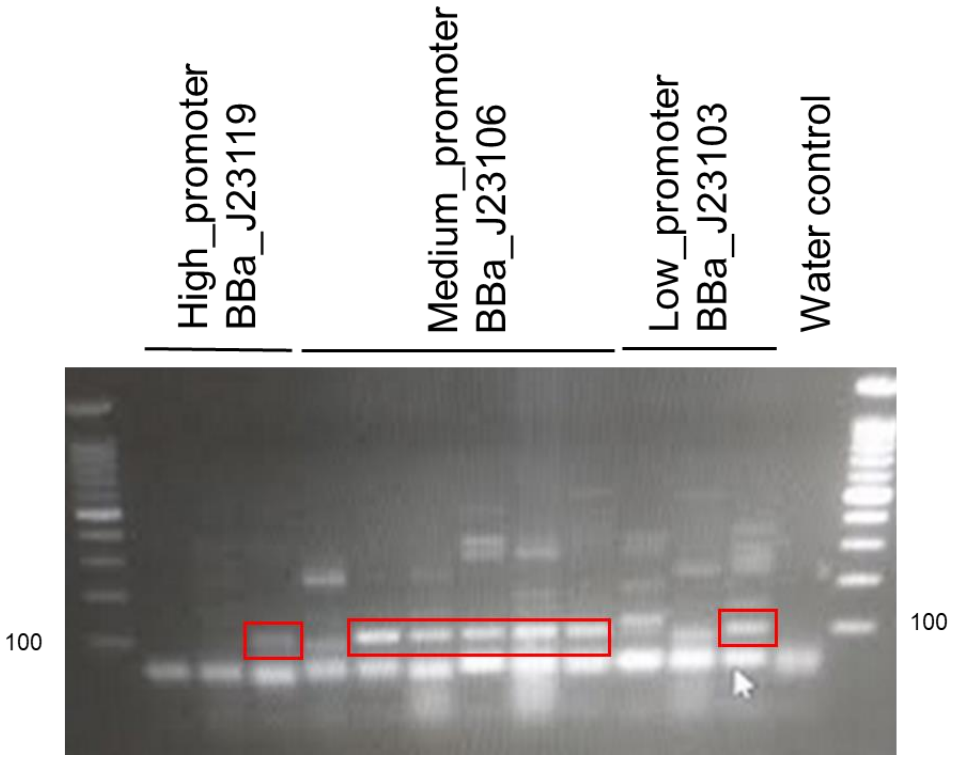


Figure 76. Colony PCR checking cloning of Anderson promoters into the pAS15a-Cm plasmid. Cloning was checked using p15a_CAT_F and promoter_R primers and successful clones produced a band that was around 100 bp and is circled in red. A PCR reaction with no template was used as a negative control.

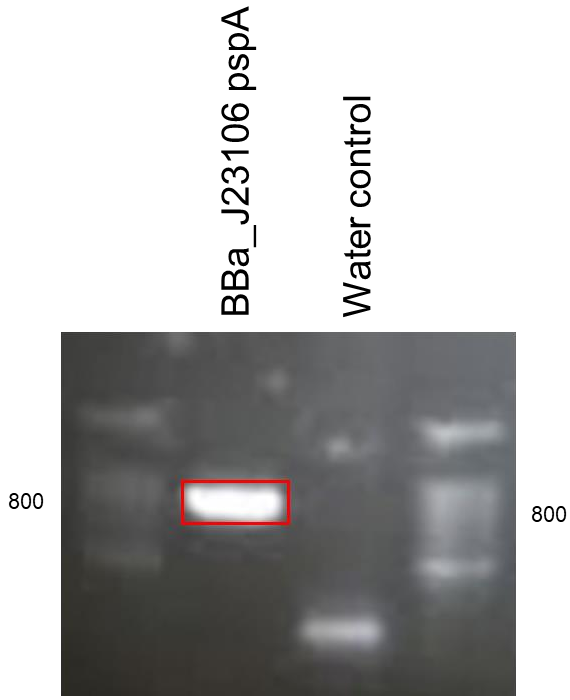


Figure 77. Colony PCR checking cloning of pspA into the pAS15a-Cm plasmid containing a medium strength Anderson promoter BBa_J23106. Cloning was checked using p15a_CAT_F and p15a_CAT_R primers. Successful clones amplified a product around 800 bp which is circled in red. A PCR reaction with no template (water) was used as a negative control. Error bars represent standard deviation.

4.3.3.1.2 Verifying protein overexpression.

Overexpression of PspA was verified using western blot, however, overexpression of other overexpressed proteins could not be verified as no antibodies against them were available and expression was not high enough to observe bands on SDS-PAGE. Bacteria were grown using the same media as in high cell density cultures (in shake flasks) as culture conditions can pay an important role in Anderson promoter activity [309]. PspA was present in all samples including the empty vector control (pAS15a_cm), which is expected as the wildtype strain contains the *pspA* gene. Expression of PspA increased significantly in cells containing the plasmid with a high inducible promoter compared to the control as seen by the significant increase in bands intensity (one-way ANOVA, $p<0.05$, N=2) (Figure 78). However, there was no significant difference in the intensity of bands representing PspA when comparing expression of medium and low constitutive promoters compared to the wildtype control (Figure 78), which may be due to PspA overexpression being too low in

these samples to detect via western blot. qPCR was also conducted to verify PspA transcription. Levels of *pspA* transcripts were proportional to the strength of the promoter used (Figure 79). The abundance of *pspA* transcripts was however lower than the empty vector control (pAS15a_cm), when co-expressing PspA with medium and low promoters which could be due to changes in cell physiology and gene regulation. qPCR was however only conducted using one biological replicate due to time limitations and further replicates would be needed to confirm these observations.

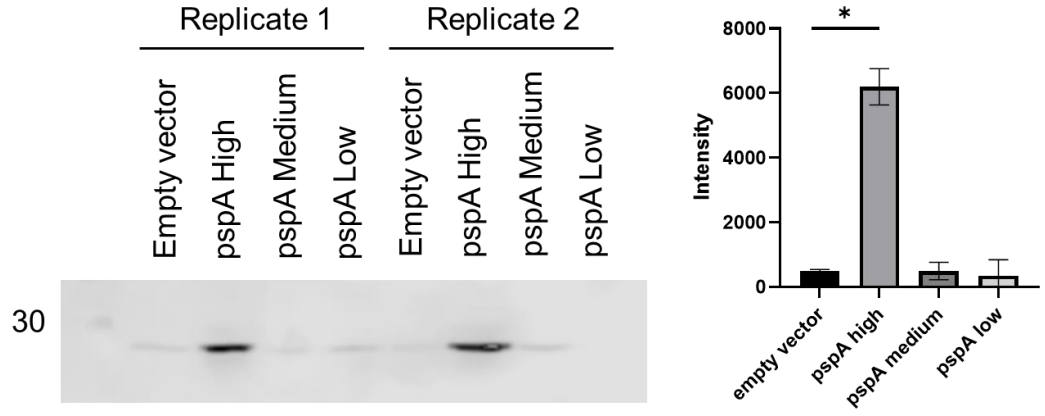


Figure 78. Western blot of E. coli measuring PspA levels when constitutively overexpressing PspA with different strength Anderson promoters (high, medium and low) compared to control containing an empty vector plasmid (pAS15a_cm). PspA has a molecular weight of around 25 kDa. E. coli was grown in shake flask conditions, using the same media used in fermenters. Biological duplicates were taken during exponential growth (5 hours). Samples were run on an SDS-PAGE to check for equal loading. Band Intensity was measured using ImageJ and analysed using a one-way ANOVA. Differentially expressed genes ($p<0.05$) are indicated by a star. Error bars represent standard deviation.

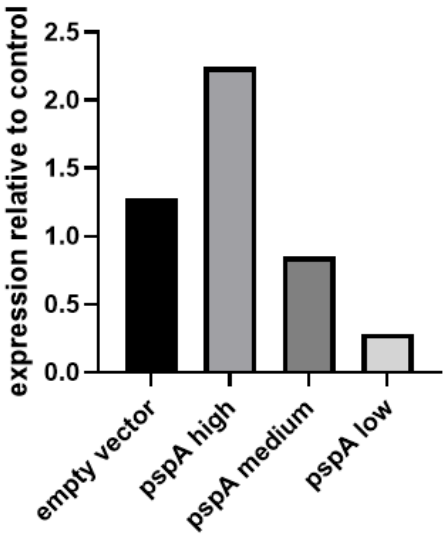


Figure 79. qPCR analysis of pspA expression in strains overexpression High, medium and low levels of pspA expression compared to an empty vector control (pAS15a_cm). PspA expression was quantified relative to an internal standard (rpoA). E. coli was grown in shake flask cultures, using the same media used in fermenters and samples were taken during exponential growth (5 hours). Technical replicates and a no reverse transcriptase and water control were also conducted. Figure represents one biological replicate.

4.3.3.1.3 Effect of protein overexpression on growth

Plasmid presence and protein overexpression can have a detrimental effect on growth which would not be desirable when developing a novel overexpression strain [150], [152], [154], [310], [311]. Growth of strains was therefore compared to wildtype and empty vector (pAS15a-cm) controls. Bacteria were grown in similar media to that used in FDB fermenters in 96 well plates to allow higher throughput of experiments and were incubated with shaking at 37 °C (see methods section 2.2.1). Bacteria did not grow to as high densities as fermenters probably due to limited aeration in 96 well plates and differences in path length in 96 well plates (Figure 80) [289]. Presence of the empty vector pAS15a-Cm had no effect on growth compared to wildtype control. Furthermore, none of the proteins that were expressed had a negative impact on growth compared to the wildtype and empty vector controls (Figure 80).

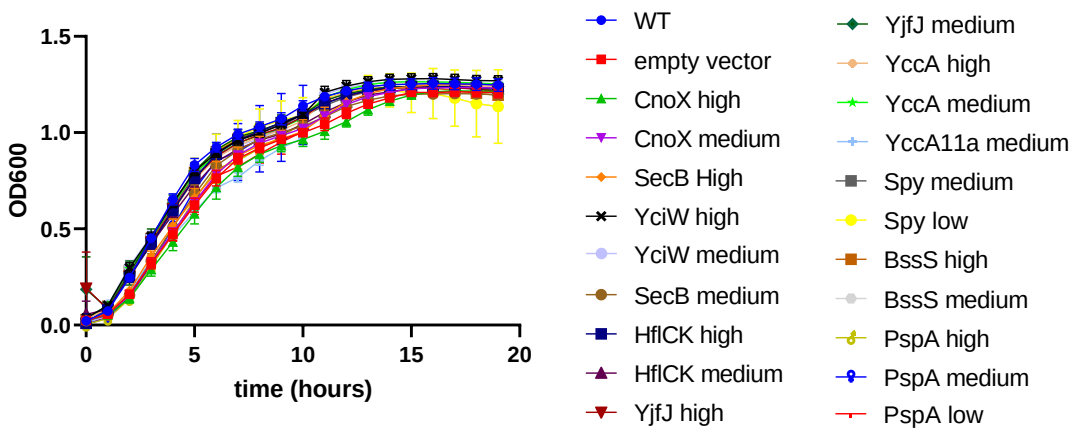


Figure 80. Growth curves comparing growth of E. coli when co-expressing genes compared to a wildtype (WT) and empty vector plasmid (pAS15a-Cm). **Bacteria were grown at 37 °C in a 96 well plate in media similar to that used previously in the Ambr fermenters (yeast extract, M9 trace salts and glycerol). OD_{600} was measured for three technical replicates and three biological replicates. The average optical density across each technical replicate was used to determine OD_{600}. Error bars represent one standard deviation across biological replicates.**

4.3.3.2 Knocking out genes

4.3.3.2.1 Lambda red recombineering

Genes were knocked out using lambda red recombineering [237]. Other methods of creating knockouts were also trialled to reduce the effects of gratuitous overexpression of lambda red genes (pDoc) [308] or create seamless mutations (pKO3) [312] which is less likely to affect expression of downstream genes and does not require an antibiotic resistance selection marker [18]. However, these methods were either time consuming or did not work in my hands.

Results from creating all 10 knockouts will not be covered in this section, however the process of creating a *pspA* knockout will be used to illustrate this process. PCR was used to amplify products containing the antibiotic resistance gene flanked with up to 50 bp upstream and 50 bp downstream of the gene of interest as can be seen by the presence of a 1500 bp band (Figure 81). In most knockouts, a kanamycin resistance gene (1500 bp product) was amplified from pKD4 (if the gene was part of an operon) or pKD13. pKD42 was used to amplify a

chloramphenicol resistance gene (1000 bp) when knocking out *cpxAR* in the Δ*pspA* Δ*cpxAR* double knockout.

These PCR products were purified and transformed into cells overexpressing lambda red recombineering genes from pKD46 that were grown at 30 °C. Successful clones were then selected on antibiotic plates (in this case Kanamycin). Bacteria were grown at grown at 37-42 °C and removal of the temperature sensitive lambda red plasmid was verified by streaking colonies onto ampicillin plates. Colony PCR was then used to verify knockouts using primers that flanked the knocked out region and produce a 1500 bp band seen in Figure 82A and primer pairs that flank the region and overlap the antibiotic resistance cassette producing a 1000 bp band that is only amplified in the knockout (Figure 82B). PCR products were purified and sequenced to further confirm knockouts.

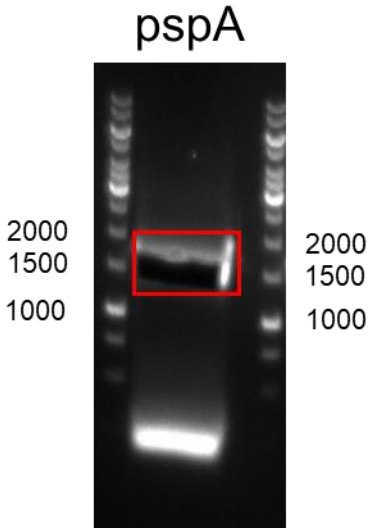


Figure 81. DNA gel of PCR used to make pspA knockout and amplify kanR from pKD4. The PCR product was amplified using pspA_knockout_F and pspA_knockout_R. The expected size of the product is around 1500 bp. This band is surrounded by red and has been cut out of the gel.

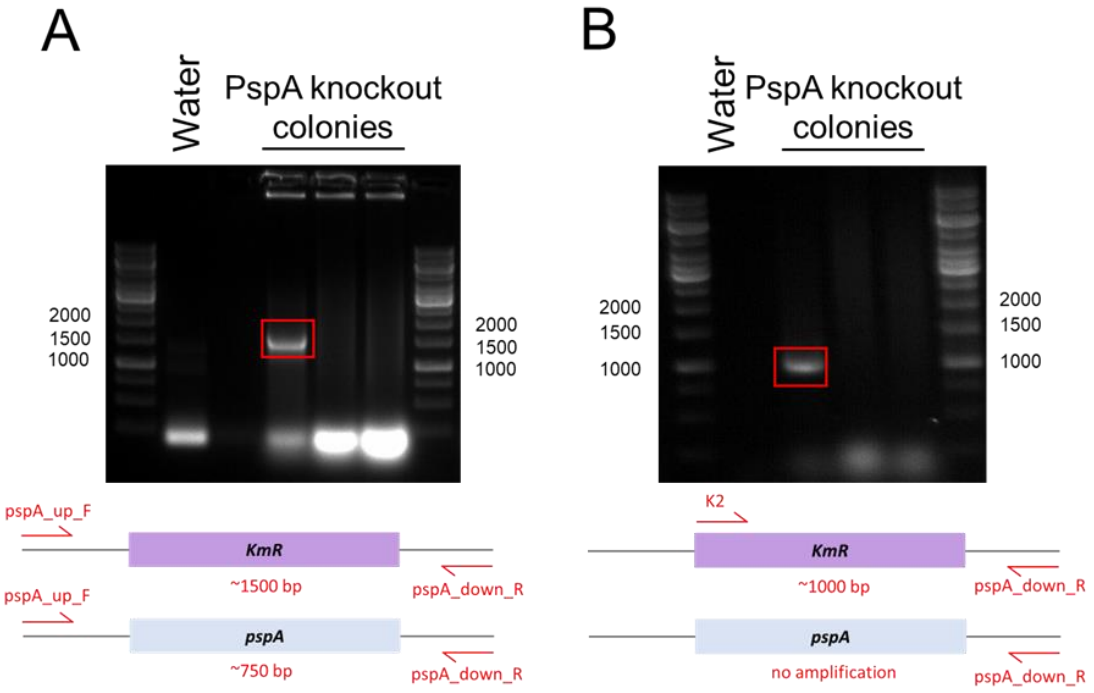


Figure 82. Colony PCR used to check pspA knockouts. Colonies were checked using two primer sets. One set of primers flanked the region that was knocked out (pspF_up_F and pspA_down_R) and successful knockouts lead to the amplification of 1500 bp products circled in red. Colonies were also checked using primers that bound to the kanamycin resistance gene and downstream of pspA (K2 and pspA_down_R). Successful knockouts lead to the amplification of a 1000 bp product which is surrounded in red. Samples containing no template (water) were used as a negative control.

4.3.3.2.2 Effect of knocking out genes on growth

Knocking out genes can have a detrimental effect on growth which would not be desirable when developing a overexpression strains [297], [298]. Growth of knockout strains and strains that combined mutations was therefore tested when grown in media similar to that used in Ambr fermenters in a 96-well plate at 37 °C (see methods section 2.2.1). None of the knockout strains had a negative impact on growth compared to the wildtype (Figure 83) and were therefore used for the rest of this study.

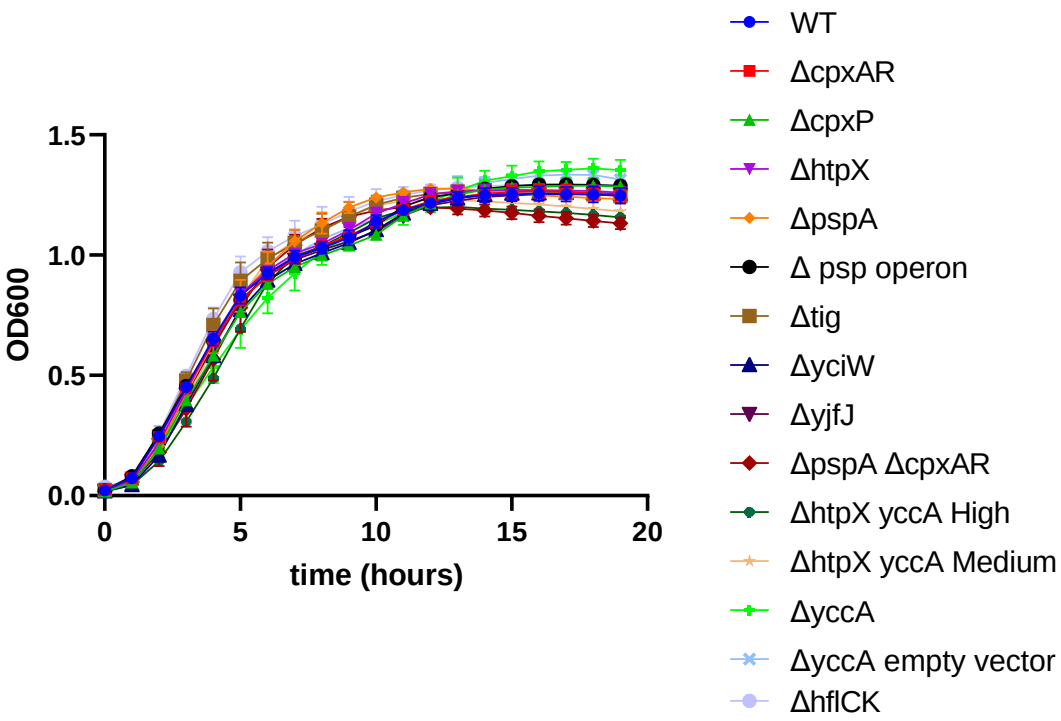


Figure 83. Growth curves comparing growth of E. coli when knocking out genes compared to a wildtype (WT) control. Strains that combined knocking out and co-expressing genes were also tested. **Bacteria were grown at 37 °C in a 96 well plate in similar media than that used in Ambr fermenters (yeast extract, M9 trace salts and glycerol). Three technical replicates and three biological replicates were taken for each sample. The average optical density across each technical replicate was used to determine OD_{600}. Error bars represent one standard deviation across biological replicates.**

4.3.3.3 Optimizing amino acid addition

ScFv secretion had a negative impact on expression of multiple genes involved in amino acid synthesis including an increase in the CysB regulon (positive regulator of cysteine biosynthesis) and the MetJ regulon (negative regulator of methionine biosynthesis). Therefore cysteine, methionine and casamino acids (mixture of amino acids) were added to media to test if this improved scFv production and secretion.

High concentrations of amino acid can be toxic to *E. coli* [313], [314]. Bacterial growth was therefore tested with different concentrations of amino acids. The concentrations of amino acids used was determined based on

literature that indicated that 0.2 mM of cysteine was toxic and 4.2 mM methionine was toxic to *E. coli* cells [313]–[315].

Addition of up to 2.5 mM cysteine or methionine had no significant impact, negative or positive on *E. coli* growth (Figure 84). Therefore 2.5 mM Cysteine and 2.5 mM methionine additions were used in all future experiments as this was the highest concentration of amino acid addition tested that had no negative impact on growth. Surprisingly however, addition of 2% casamino acids lead to a significant increase in growth of *E. coli*, as seen by the significant difference in optical density between 8-13 hours (two-way ANOVA, $p<0.05$, N=3). This indicates that growth media may be lacking amino acids. There was however no significant difference in growth when adding 1% and 0.5% casamino acids. 2% casamino acid concentrations were therefore used to test whether amino acid addition improved scFv expression and secretion.

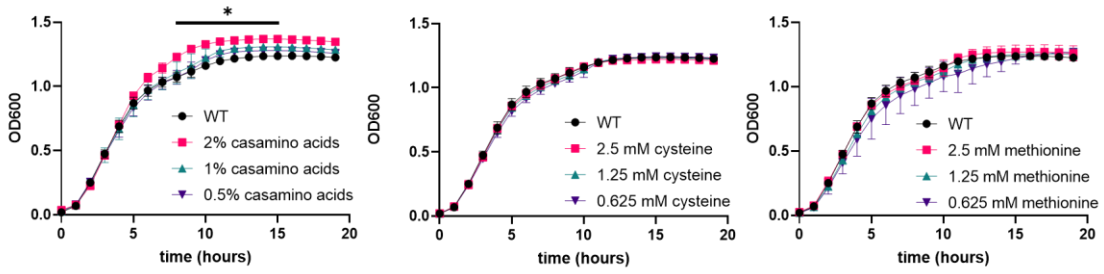


Figure 84. Growth curves comparing growth of wildtype E. coli* with different concentrations of casamino acids, cysteine, and methionine added to media compared to no amino acid addition (WT). Bacteria were grown at 37 °C in a 96 well plate in media containing yeast extract, trace salts and glycerol. OD_{600} was measured for three technical replicates and three biological replicates. The average optical density across each technical replicate was used to determine OD_{600}. Error bars represent one standard deviation across biological replicates. Samples were analysed using a two-way ANOVA. *Error bars represent standard deviation.

4.3.4 Testing mutants and altered media for improved recombinant protein production and secretion

20 overexpression strains, 10 knockout strains, 3 strains that combined mutations and 3 alterations to media were tested as part of this study. In the following section, modifications were tested to determine whether they improved scFv expression and secretion under shake flask conditions. Strains

that successfully improved scFv production were then tested to see if they improved scFv production under industrially relevant High Cell Density Culture (HCDC) conditions. Finally, successful strains were tested for improved production of other recombinant periplasmic proteins (Hel4, sfGFP and HGH) under shake flask conditions. These steps have been represented in the flow chart in Figure 85.

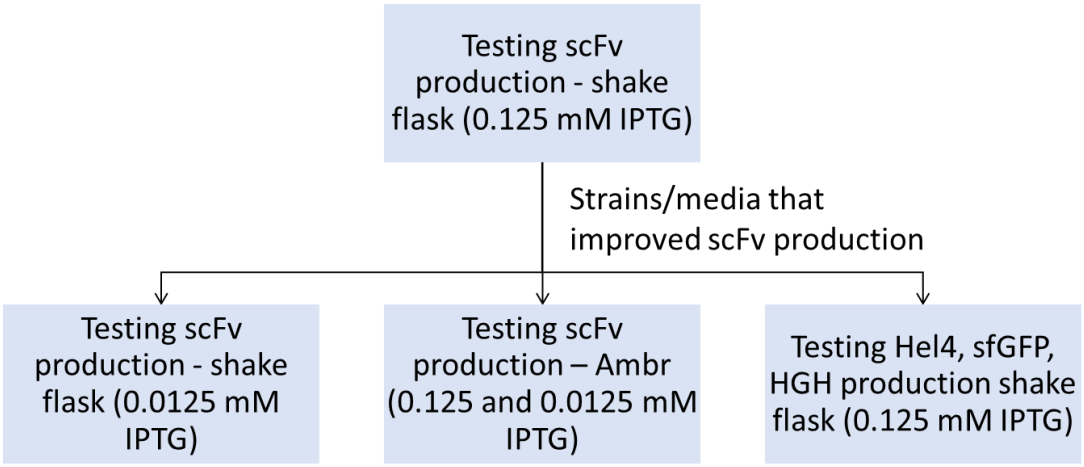


Figure 85. Flowchart illustrating the approach used to test engineered strains for enhanced scFv production and secretion.

4.3.4.1.1 Introduction

Strain and media alterations were tested to see if they improved production and secretion of the scFv by overexpressing the scFv in shake flasks. Wildtype and empty vector (pAS15a-Cm) strains were used as controls for knockout strains and overexpression strains respectively. Technical replicates were taken for all samples. ScFv expression was induced mid log phase (OD_{600} 3-4) with the high concentration of IPTG (0.125 mM) as this presented the highest level of stress on the cells and we assume overloaded the secretion system. If a strain improved production of the scFv when induced with 0.125 mM IPTG, it was also tested when inducing with a lower concentration of IPTG (0.0125 mM) that overloaded the secretion system less (Figure 85). Samples were taken 3 hours post induction, as this is a standard time point and is still during the exponential phase. Samples were also taken 24 hours post induction, but not all were analysed. Supernatant samples were also gathered to determine if cells were lysing. Periplasmic fractions were not taken routinely due to time

constraints and the fact that periplasmic fractions are not extracted as part of industrial processes at FDB.

ScFv abundance and secretion was quantified using SDS-PAGE of whole cell samples as secreted and unsecreted scFv can be differentiated based on molecular weight (the unsecreted form of the scFv contains an OmpA signal peptide). This presents the advantage of being able to rapidly determine large differences in scFv abundance. However, this approach is also challenging due to issues with dynamic range, difficulties separating the processed and unprocessed scFv bands, background proteins interfering with protein abundances and differences in protein loading in each lane. Therefore, band intensities were also determined using image analysis software (ImageJ). The background was removed using the rolling ball function in ImageJ and peak intensity was normalised to the lane intensity without the scFv bands or by comparing the intensity of the 12 kDa secreted scFv to the 14 kDa unsecreted scFv (see methods section 2.4.2.4). However, image analysis also presented several issues due to background intensity, streaky gels, insufficient separation of bands and lanes not running straight [316], [317]. It is therefore important to consider both visual analysis of SDS-PAGEs and image analysis of band intensities when determining if an alteration improves recombinant protein overexpression as well as analysing multiple replicates to ensure responses are consistent.

Ideally, scFv abundance would have been quantified using an antibody test (western blots or ELISAs), by measuring protein activity (eg. fluorescence, enzymatic activity) and/or by measuring protein abundance after purification. However, this was not possible as no antibodies or activity assays were available for the scFv and purification of the scFv would have required multiple purification steps that would have been extremely time consuming. An anti-scFv antibody (anti igkv1-5) was trialled to determine if this bound to the anti-TNF scFv, however, our scFv was not detected using this method. A C-terminal FLAG tag was added to the scFv so that it could be detected using anti-FLAG antibodies. However, overexpression of the FLAG tagged scFv greatly reduced the quantity of scFv expressed compared to when it was not tagged (Appendix 28) and

affected the growth of strains. This FLAG tagged scFv was therefore not used in this study.

Most alterations to strains and media tested in this study (altering expression of the envelope stress response genes, periplasmic and cytoplasmic chaperones, flagella, genes of unknown function and amino acid additions) did not significantly improve production and secretion of the scFv. These strains and alterations to media were therefore not tested to determine if they improved recombinant protein secretion under industrially relevant conditions or improved secretion of other recombinant proteins. Due to the large number of strains tested the data will not be described in detail here, however, all gels can be found in Appendix 29, Appendix 30 and Appendix 31 and image analysis of scFv intensity can be found in Figure 86. There seemed to be a non-statistically significant increase in scFv production when co-expressing high and medium quantities of Spy and medium quantities of SecB based on image analysis (Figure 86), however, no differences were observable by eye (Appendix 29) and were therefore probably due to inherent flaws with image analysis described previously.

To our knowledge, most strain and media alterations tested in this study had not been previously reported on in the literature. However, overexpressing PspA had previously been shown to improve recombinant protein secretion [146]–[148]. The fact that overexpressing PspA did not improve scFv production in this study could be due to differences in culture conditions, properties of overexpressed recombinant proteins, the quantity of PspA produced and whether the secretion system was overloaded. Studies also found that knocking out Tig improved secretion of leech carboxypeptidase inhibitor (LCI) to the periplasm [131]. However, knocking out Tig did not improve scFv production and secretion in this study, which may be due to different properties of specific overexpressed recombinant proteins, culture conditions and different secretion pathways used (LCI was secreted co-translationally whereas the scFv was secreted post-translationally). Overexpression of genes involved in cysteine biosynthesis (CysK [216]) and overexpression of amino acid uptake pathways (PrsA [201]) has previously been shown to improve the overall expression of certain recombinant proteins, however, in this study, adition of cysteine,

methionine and casamino acids to media did not improve overall production and secretion of the scFv. This could be due to multiple factors including differences in the amino acid content of the protein that is being overexpressed, the amino acid content of the media and whether the amino acids that are being added to media are being used by the cell. Studies have also tested the effect of overexpressing SecB, but did not see any improvements in recombinant protein production and secretion, as was also the case in this study [101].

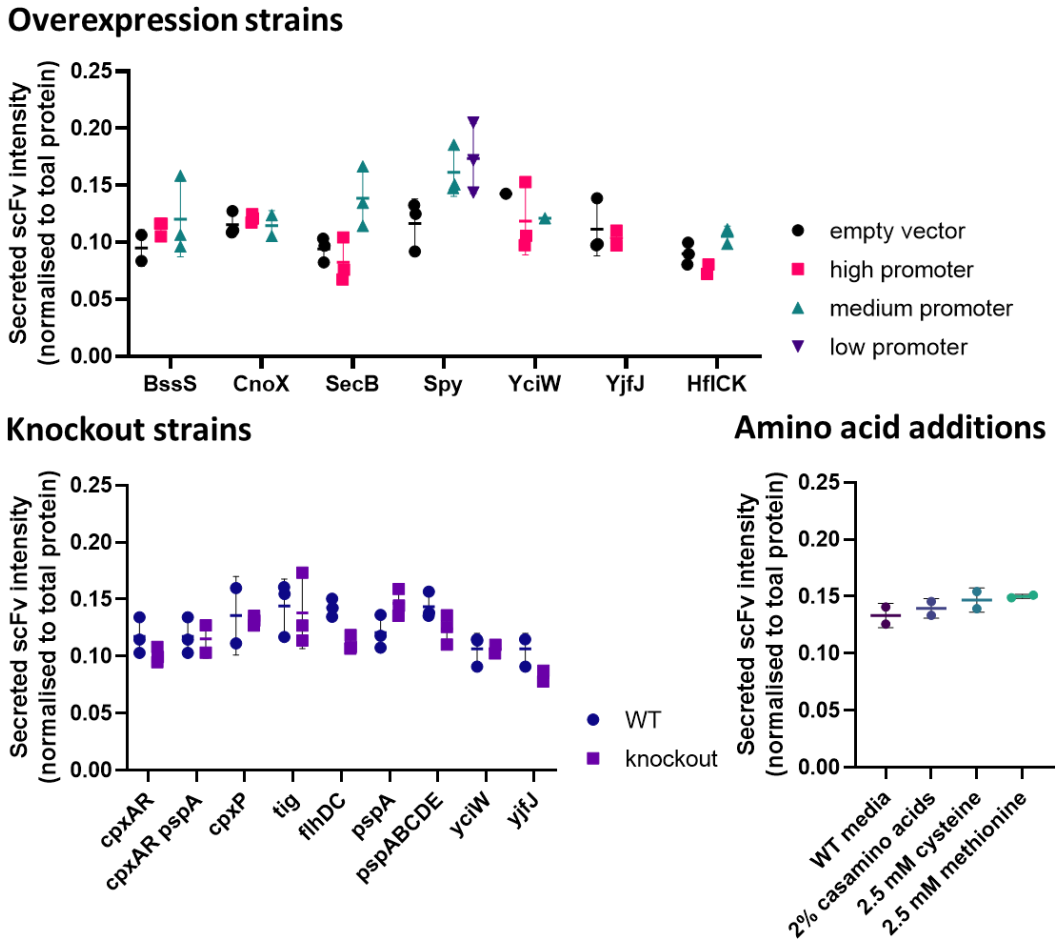


Figure 86. Intensity of secreted scFv bands in overexpression strains, knockouts and alterations to medium. Band intensity was measured using ImageJ analysis of SDS-PAGE gels and was normalised to lane intensity. Wild-type (WT) strains and media and an empty vector strain (pAS15a_cm) was used as a control. Technical triplicates were taken for each sample other than when issue arose with sampling or bands were not quantifiable as they were smeary. Error bars represent standard deviation.

There were however some strain alterations that reduced SecY and membrane protein degradation (Δ*htpX,* YccA and YccA11 overexpression) which

were shown to improve scFv secretion and will be discussed in more depth in the following sections.

4.3.4.1.2 Characterisation of scFv production in the Δ*htpX* strain

HtpX is a membrane protease that is known to degrade membrane proteins including SecY [279] that was upregulated at the transcript and protein level upon scFv secretion (see section 3.3.3.5.2.2). It was hypothesised that inhibiting membrane protein degradation would improve scFv secretion by alleviating stress and increasing the Sec secretion systems capabilities. A *htpX* knockout strain was therefore grown in shake flask conditions while overexpressing the scFv in the periplasm to test whether it outperformed the wildtype strain. The scFv was overexpressed using the P11-scFv+sp strain and induced with 0.125 mM IPTG mid log phase. Samples were taken 3 hours post induction and whole cells were analysed using SDS-PAGE to quantify scFv abundance and secretion. Two of the three biological replicates (replicate 1 and 3) improved secretion of the scFv to the periplasm in the Δ*htpX* strain was observed, as indicated by the increased ratio of secreted (12 kDa) to non-secreted (14 kDa) scFv (on average a 90% increase for replicate 1 and a 34% increase for replicate 2) (Figure 21). For replicate 3 the difference was statistically significant based on image analysis data (t-test, N=3) (Figure 87). For replicates 1 and 2 there was also an increase in the intensity of secreted scFv bands in cultures where *htpX* had been knocked out, (observable by eye and using image analysis), however, this difference was small (an average 26% improvement) and not statistically significant (Figure 87).

The *htpX* deletion strain was also tested for improved scFv production when inducing recombinant protein expression with lower concentrations of IPTG (0.0125 mM). Under these conditions only secreted scFv (12 kDa) was detected in both parental and Δ*htpX* strains (Figure 88). Thus, it appeared that the Sec secretion system had sufficient capacity to translocate scFv under these low induction conditions and therefore enhancing the stability/levels of SecY by eliminating *htpX* had no beneficial effect. Furthermore, based on previously generated proteomic data, HtpX expression only increased in response to high levels of scFv expression (0.125 mM IPTG), but not when inducing low levels of

scFv expression (0.0125 mM IPTG) (see section 3.3.3.5.2.2). This indicated that lower levels of scFv expression did not affect SecY and membrane protein degradation pathways, which could therefore explain why altering expression of these pathways was not beneficial under these conditions.

To our knowledge, no other studies had examined the effect of knocking out *htpX* on expression of secreted recombinant proteins.

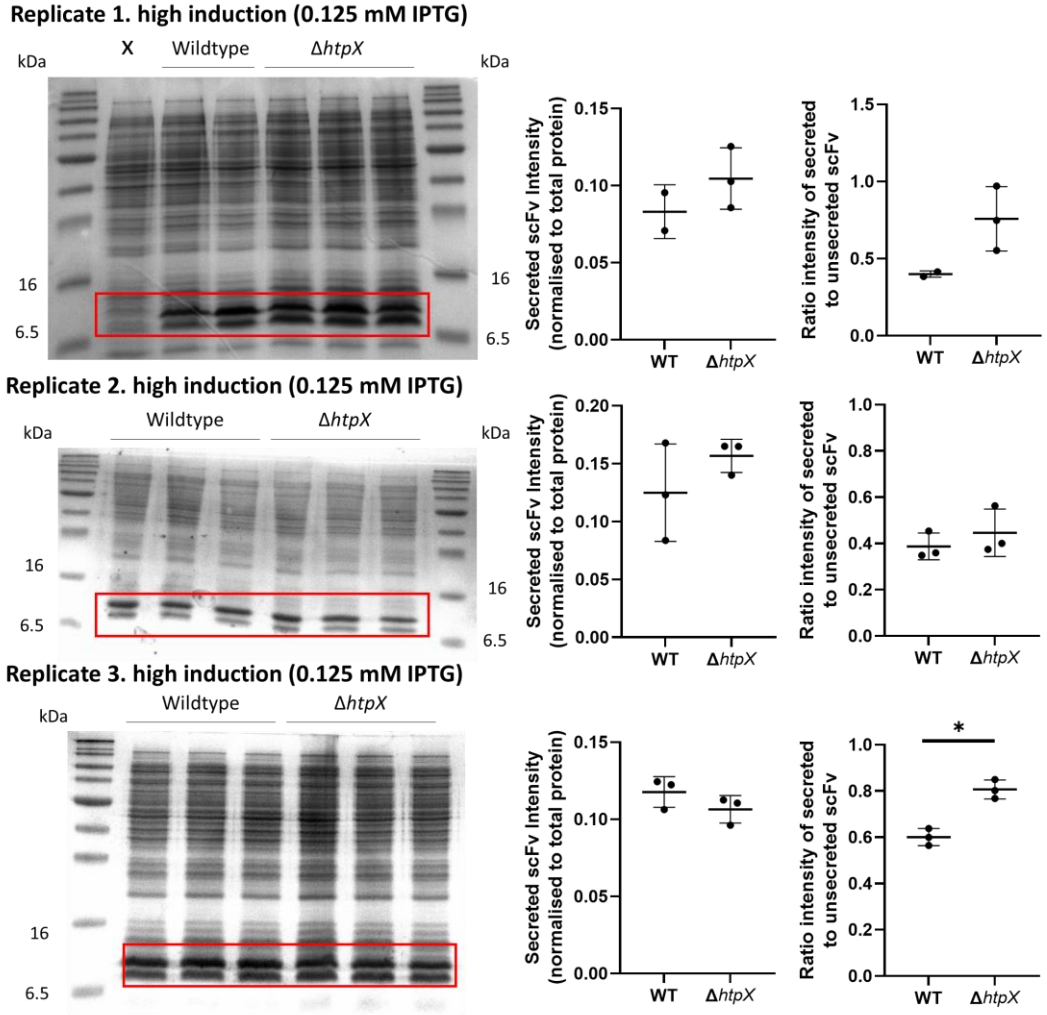


Figure 87. SDS-PAGE analysis of knocking out htpX (ΔhtpX) on scFv expression and secretion compared to wild-type control (WT) when induced with high concentrations of IPTG (0.125 mM) under shake flask conditions. Sample loading was normalised to bacterial concentration (OD600). The scFv (indicated within the red box) has a predicted molecular weight of 14 kDa with the signal peptide and 12 kDa without (secreted form). Technical triplicates were separated on the same gel. Abundance of the secreted scFv (12 kDa band) was determined using densitometry using ImageJ and was normalised to the intensity of each lane or the intensity of the non-secreted 14 kDa upper band. Statistical analysis was conducted using GraphPad Prism and samples were analysed by a standard t-test and differences deemed statistically significant are indicated (star). Error bars represent standard deviation.

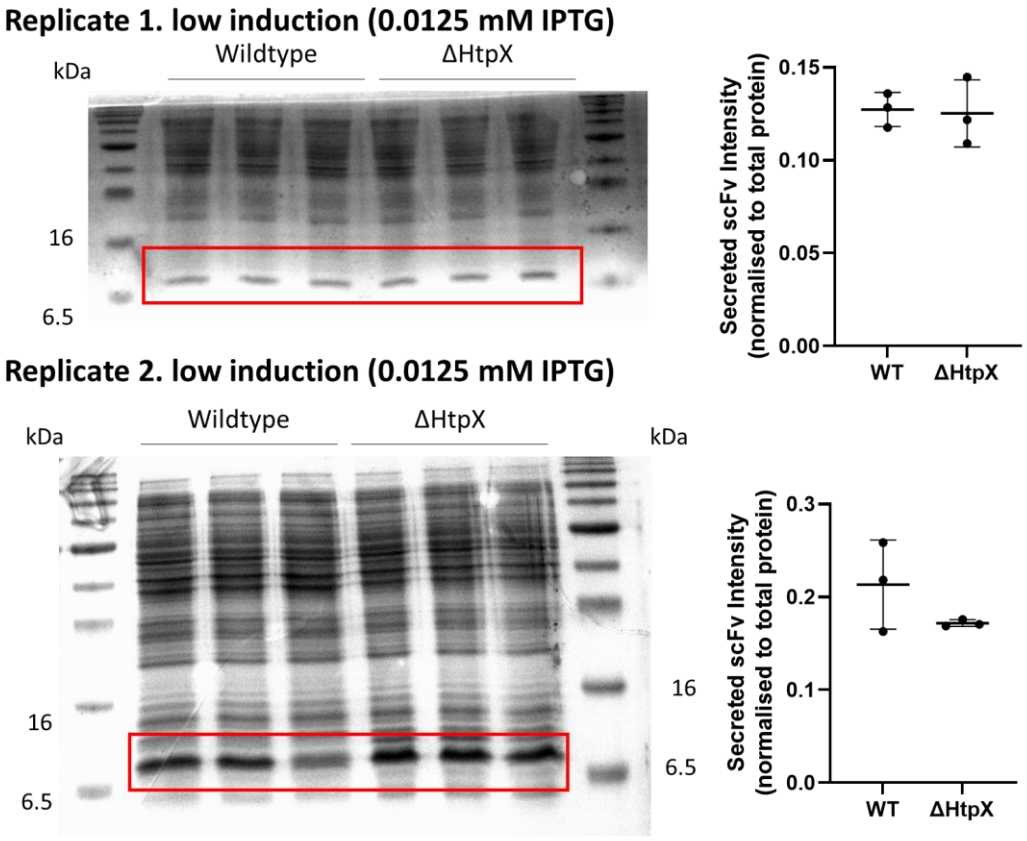


Figure 88. SDS-PAGE analysis of knocking out htpX (ΔhtpX) on scFv expression and secretion compared to wild-type control (WT) when induced with low concentrations of IPTG (0.0125 mM) under shake flask conditions. Sample loading was normalised to bacterial concentration (OD600). The secreted scFv (within the red box) has a predicted molecular weight of 12 kDa; note that in contrast to high-level induction unsecreted (14 kDa) scFv was not detected. Technical repeats were separated on the same gel and each gel represents a biological replicate. Abundance of the secreted scFv band (12 kDa) was determined by densitometry using ImageJ and was normalised to the intensity of each lane. Statistical analysis was conducted using GraphPad Prism and samples were analysed using a standard t-test. Error bars represent standard deviation.

4.3.4.1.3 Characterisation of scFv production in YccA and YccA11 overexpressing strains

YccA is a negative inhibitor of the membrane protease FtsH which is known to degrade SecY and membrane proteins [281], [318] and was downregulated upon scFv secretion (see section 3.3.3.5.2.2). It was therefore hypothesised that overexpressing YccA would improve scFv secretion by increasing the capacity of the secretion system and alleviating stress.

The parental scFv production strain was transformed with *yccA* expression plasmids. The *yccA* gene was expressed from either a medium or high strength constitutive promoter. An increase in the intensity of bands associated with the secreted scFv (12 kDa), particularly for high levels of YccA expression was observed. There was a significant increase in the intensity of the secreted scFv band (12 kDa) when high levels of YccA were expressed in one of the replicates led to a 77% increase in band intensity compared to the control (replicate 2) (one way ANOVA, $p<0.05$, $n=3$) (Figure 89). There was however no difference in the ratio of secreted (12 kDa) and unsecreted (14 kDa) scFv bands when overexpressing YccA compared to empty vector controls (pAS15a_cm).

The effect of YccA overexpression was also tested when inducing lower levels of scFv expression (0.0125 mM IPTG), that were previously identified as being within the capacity of the Sec system (Figure 22). Co-expression of YccA from a medium strength promoter had a negative effect on scFv production and secretion as seen by the significant decrease in scFv band intensity compared to the empty vector control (Replicate 1, one-way ANOVA, $p<0.05$ $n=3$) (Figure 89). This could be explained by the fact that the secretion system was not overloaded under these conditions and therefore degradation of SecY and other membrane proteins was not detrimental. Moreover, YccA co-expression may be using up important cellular resources (YccA is a membrane protein and therefore needs to be secreted) and/or may be affecting expression of detrimental stress response pathways.

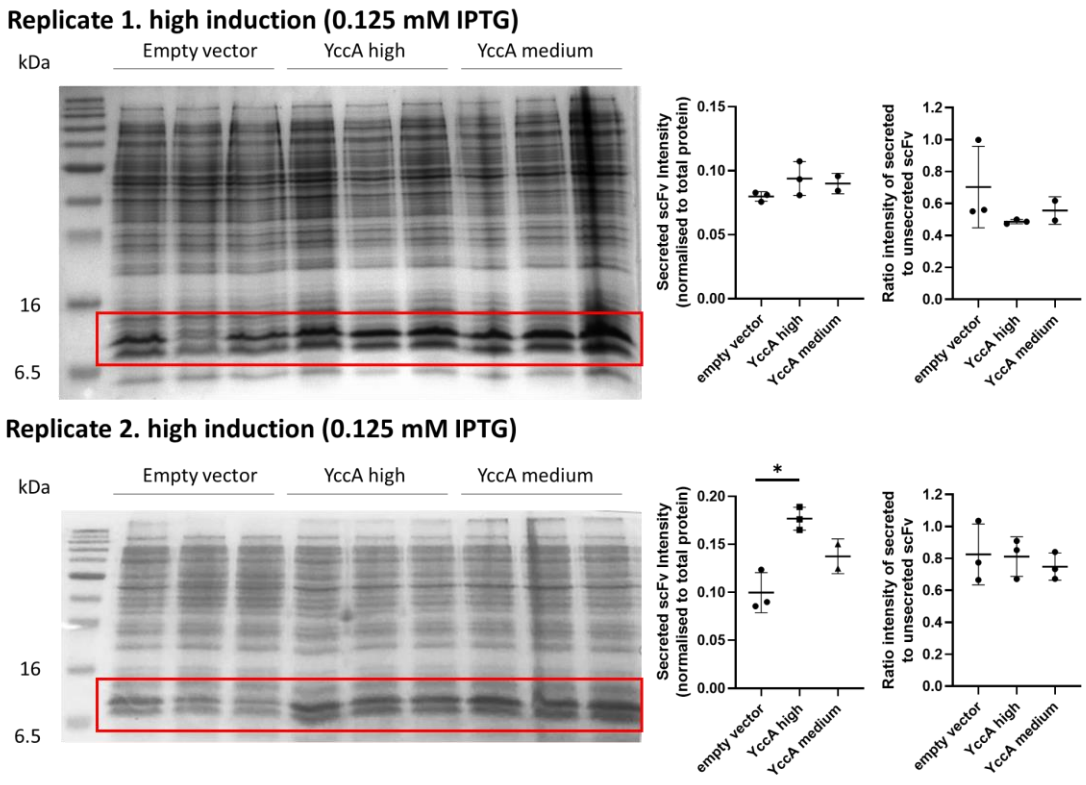


Figure 89. SDS-PAGE analysis of overexpressing different levels of YccA (high and medium constitutive promoter) on scFv expression and secretion, compared to an empty vector control (pAS15a_cm). The scFv was induced with a high concentration of IPTG (0.125 mM) under shake flask conditions. Sample loading was normalised to bacterial concentration (OD600). The scFv (within the red box) has a predicted molecular weight of 14 kDa with the signal peptide and 12 kDa without (secreted form). Technical repeats were analysed on the same gel and each gel represents a biological replicate. Abundance of the secreted scFv band (12 kDa) was determined by densitometry using ImageJ and was normalised to the intensity of each lane. Lanes containing smears were discarded from densitometry analysis (replicate 1, lane 10 and replicate 2, lane 9). Statistical analysis was conducted using GraphPad Prism using a one-way ANOVA and significant differences are indicated by a star. Error bars represent standard deviation.

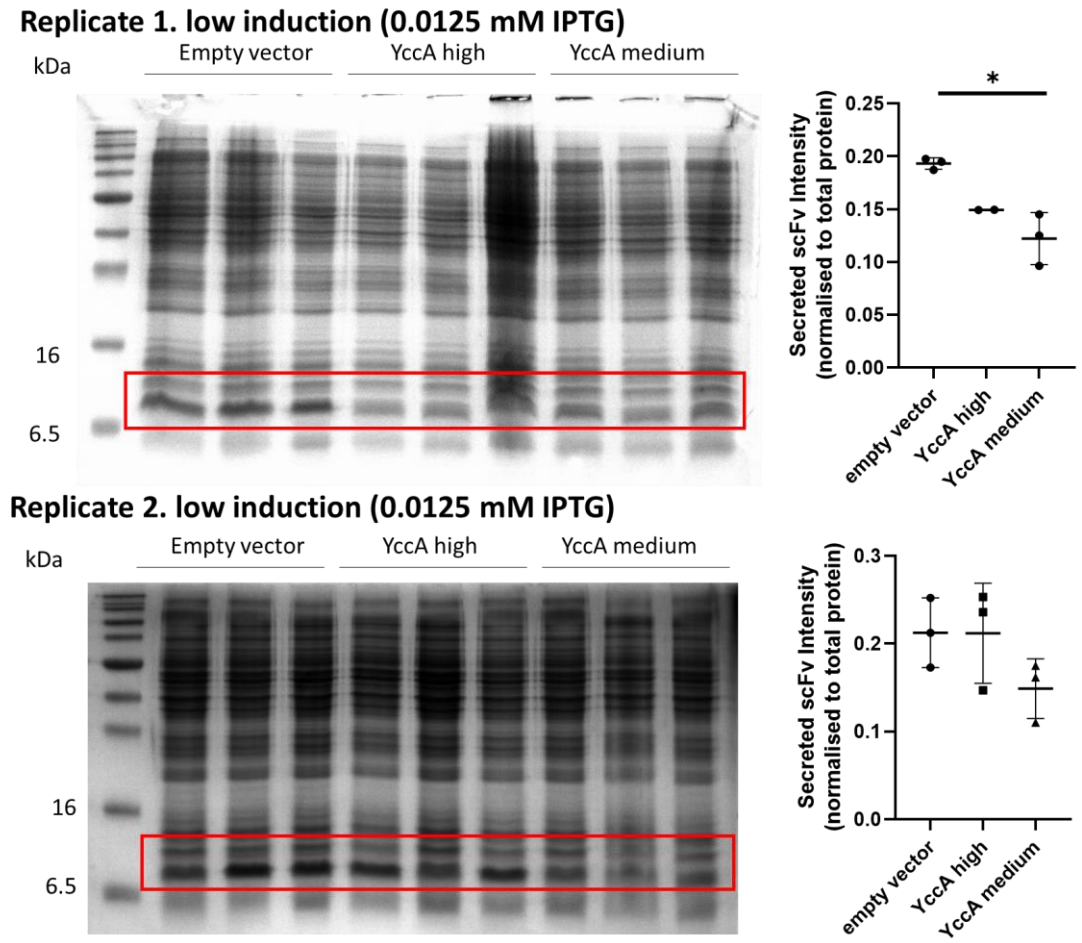


Figure 90. SDS-PAGE analysis of overexpressing different levels of* YccA *(high and medium constitutive promoter) on scFv expression and secretion, compared to an empty vector control (pAS15a_cm). The scFv was induced with low concentrations of IPTG (0.0125 mM) under shake flask conditions. The scFv (indicated within the red box) has a predicted molecular weight of 12 kDa without the signal peptide (secreted form). Technical repeats were analysed on the same gel and each gel represents a biological replicate. Abundance of the secreted scFv band (12 kDa) was determined by densitometry using ImageJ and was normalised to the intensity of each lane. Lanes containing smears were discarded from densitometry analysis (replicate 1, lane 7). Statistical analysis was conducted using GraphPad Prism using a one-way ANOVA and significant differences are indicated by a star. Error bars represent standard deviation.

The observation that YccA co-expression improved production of the scFv when the Sec system was overloaded, prompted experiments in which YccA11, a truncated form of YccA that is more active, was overexpressed [281], [299], [300]. Plasmids were constructed that expressed YccA11 from a medium strength constitutive promoter as there was not sufficient time to also co-express YccA11 with a high promoter. Co-expressing medium levels of YccA11

led to an average 26% and 55% increase in the intensity of the secreted scFv band and a 5% and 7% increase in the ratio of secreted (12 kDa) to unsecreted (14 kDa) scFv when compared to the empty vector control and the strain expressing wild-type YccA respectively (Figure 91). This difference was however not deemed statistically significant and surprisingly no difference was seen between scFv expression in the empty vector control and strains overexpressing wild-type YccA. This may however be due to poor band resolution in these gels and further replicates would need to be conducted to confirm this.

The experiments described above were conducted in strains containing a wild-type chromosomal *yccA* gene, which could potentially confound any beneficial effects of YccA11 expression. Therefore, the effects of YccA11 co-expression on scFv production were tested in a *yccA* knockout strain. Overexpression of YccA11 in the Δ*yccA* background led to an average 1.9% increase in the intensity of the secreted scFv band (12 kDa) and a 3.4% in the ratio of secreted (12 kDa) to unsecreted (14 kDa) scFv, compared to an empty vector control (Figure 92). This difference was not statistically significant, even though the ratio of secreted and unsecreted scFv appeared to increase when observing gels by eye. Overall, these experiments suggest that overexpressing the mutant form of YccA (YccA11) improves scFv secretion, however further replicates would be needed to support this suggestion. The *ΔyccA*, YccA11 medium strain was only generated towards the end of the project and was therefore not used in the rest of this study.

Previous studies have observed that expressing YccA and YccA11 improved secretion of the outer membrane protein LamB-LacZ and lowered secretion system overloading [280] which is consistent with observations in this study that overexpressing YccA and YccA11 improves production of the recombinant periplasmic protein used in this (scFv) when the secretion system was overloaded (0.125 mM IPTG induction).

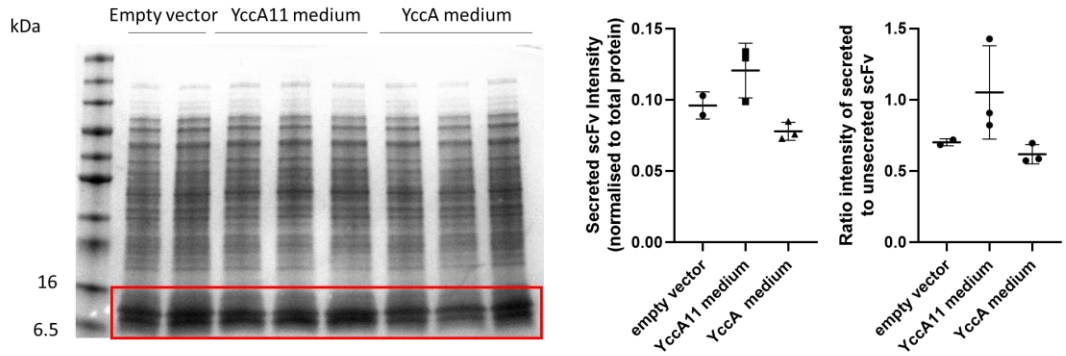


Figure 91. SDS-PAGE analysis of the effect of overexpressing a YccA variant (YccA11) on scFv production. The scFv expression strain was transformed with a plasmid expressing* YccA11 *under the control of a medium strength constitutive promoter. For comparison control strains were transformed with an empty vector (pAS15a_cm) or a plasmid overexpressing wild-type* YccA *from the same medium strength promoter. The scFv was induced with high concentrations of IPTG (0.125 mM) under shake flask conditions. Sample loading was normalised to bacterial concentration (OD600). The scFv (indicated within the red box) has a predicted molecular weight of 14 kDa with the signal peptide and 12 kDa without the signal peptide (secreted form). Technical repeats were separated in triplicate. Abundance of the secreted scFv band (12 kDa) was determined by densitometry using ImageJ and was normalised to the intensity of each lane. Statistical analysis was conducted in GraphPad Prism using a one-way ANOVA. Error bars represent standard deviation.

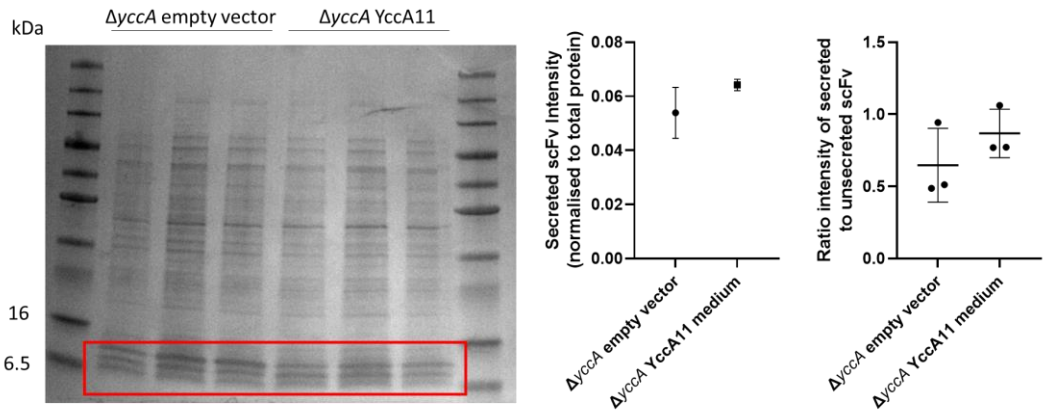


Figure 92. SDS-PAGE analysis of cultures overexpressing a YccA variant (YccA11) in the absence of wild-type yccA on scFv production. The ΔyccA strain was transformed with the scFv expression plasmid and either an empty vector (pAS15a_cm control) or a plasmid expressing* yccA11 *under the control of a medium strength constitutive promoter. The scFv was induced with high concentrations of IPTG (0.125 mM) under shake flask conditions. Sample loading was normalised to bacterial concentration (OD600). The scFv (indicated by the red box) has a predicted molecular weight of 14 kDa with the signal peptide and 12 kDa without the signal peptide (secreted form). Technical repeats were analysed in triplicate. Abundance of the secreted scFv band (12 kDa) was determined by densitometry using ImageJ and was normalised to the intensity of each lane. Statistical analysis was conducted in GraphPad Prism and samples were analysed using a t-test. Error bars represent standard deviation.

4.3.4.1.4 Combining YccA overexpression and *htpX* deletion

The effect of combining mutations by knocking out *htpX* and co-expressing YccA was also tested to determine whether further improvements in production and secretion of the scFv could be achieved. However, combining these mutations resulted in no improvement in the intensity of 12 kDa bands that represented the secreted scFv compared to knocking out *htpX* alone and led to a decrease in the ratio of secreted (12 kDa) to unsecreted (14 kDa) scFv as more of the antibody fragment remained in the cytoplasm (Figure 93). These strains were therefore not tested further.

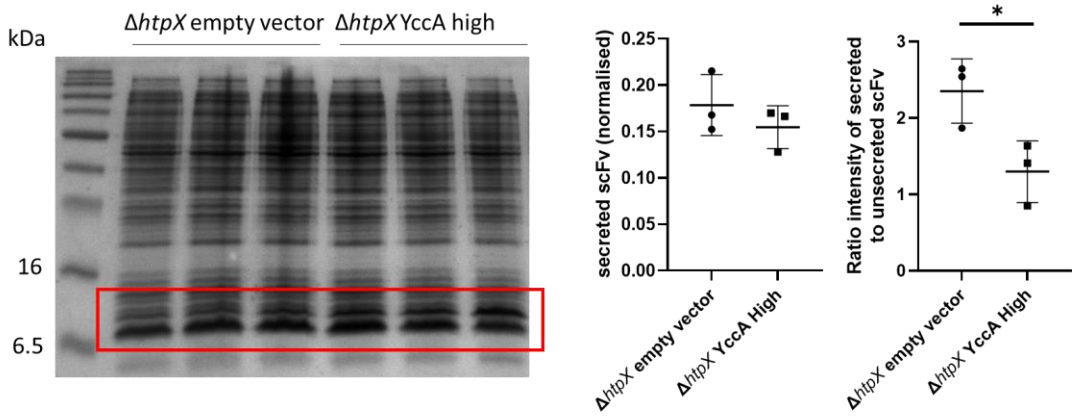


Figure 93. SDS-PAGE analysis of combining the deletion of htpX with overexpression of YccA on scFv production compared to an empty vector control (pAS15a_cm). The scFv was induced with high concentrations of IPTG (0.125 mM) under shake flask conditions. Sample loading was normalised to bacterial concentration (OD600). The scFv is surrounded in red and has a predicted molecular weight of 14 kDa with the signal peptide and 12 kDa without the signal peptide (secreted form). Technical repeats were separated in triplicate. Abundance of the secreted scFv band (12 kDa) was determined using densitometry using ImageJ and was normalised to the intensity of each lane. Statistical analysis was conducted using Graphpad Prism and samples were analysed using a t-test. Significant differences are indicated by a star (p<0.05). Error bars represent standard deviation.

4.3.4.1.5 Conclusions

Ten genes were overexpressed at two different levels, ten genes knocked out, three strains were created that combined these mutations and three medium alterations were tested for improved scFv expression and secretion in triplicate under shake flask conditions established previously (see Section 4.3.1).

Of these modifications, strains that reduced SecY and membrane protein degradation improved scFv secretion. Knocking out *htpX* improved the yields of secreted scFv and the ratio of secreted scFv compared to unsecreted scFv indicating that knocking out *htpX* reduced secretion system overloading. Overexpressing YccA improved the overall yields of scFv but not the ratio of secreted to unsecreted scFv, particularly when high constitutive promoters were used to overexpress YccA. Overexpressing a truncated form of YccA, YccA11 also further improved overall yields of the scFv and the ratio of secreted to unsecreted scFv. Combining the *htpX* knockout with YccA co-expression did not however further improve scFv secretion. Improvements were only seen when inducing scFv expression with 0.125 mM IPTG and not when inducing expression

with 0.0125 mM IPTG where the secretion system was less overloaded, which is consistent with our hypothesis that these strains reduce SecY and membrane protein degradation caused by secretion system overloading. Measuring scFv abundance using SDS-PAGE presented several disadvantages and required multiple replicates to be conducted to check that results are consistent, however, it did allow the rapid detection of very large differences in protein abundance. SDS-PAGE analysis of supernatant and periplasmic samples would have also been beneficial to further confirm that these strains improved scFv expression, however, this was limited by time constraints.

4.3.4.2 Testing successful strains – scFv production in high cell density cultures

4.3.4.2.1 Introduction

Based on previous analysis of shake flask cultures, five strains were identified that potentially improved production and secretion of the scFv coded by P11-scFv+sp, which expresses the scFv fused to an N-terminal signal peptide (sp) under the control of an IPTG-inducible promoter. All five interventions were related to reducing SecY and membrane protein degradation: (1) deletion of the *htpX* from the parental background strain; (2) overexpression of YccA from a high level promoter in the parental strain background; (3) overexpression of YccA from a medium level promoter in the parental strain background; (4) YccA11 overexpression from a medium level promoter in the parental strain background; (5) YccA11 overexpression from a medium level promoter in a Δ*yccA* strain background. However, improvements in protein yields observed in shake flask cultures are not always retained under the very different environmental conditions encountered under industrially relevant fermentation conditions (High Cell Density Cultures). Ambr bioreactors were therefore used to test the candidate strains following the same protocol used to generate OMICs samples (see methods section 2.23). Unfortunately, due to time constraints only four of these strains (1 to 4 in the list above) were tested; as the combination strain (number 5 in the list above; Δ*yccA* and YccA11 overexpression) was only generated towards the end of the project. In these HCDC experiments, scFv

production was induced with both high concentrations of IPTG (0.125 mM) that overloaded the secretion system and low concentrations of IPTG (0.0125 mM) that led to less secretion system overloading and higher levels of scFv being secreted to the periplasm. The Δ*htpX* strain was compared to a wild-type control and the YccA overexpression strains were compared to an empty vector (pAS15a-cm) control. Duplicates were generated for each sample to verify that responses were reproducible. Samples were taken 12-, 24-, 36-, and 48-hours post-induction. Two bioreactors had to be discarded due to issues with faulty pH sensors (YccA medium, P11-scFv+sp, 0.125 mM IPTG; and YccA high P11-scFv+sp, 0.0125 mM IPTG) and therefore only one sample was analysed for each of these strains. A summary of strains and conditions tested can be found in Figure 94.

Biomass production was monitored by measuring OD_{600} and dry cell weight, while SDS-PAGE of fractionated samples (whole cell, supernatant and periplasmic fractions) was used to measure scFv production and secretion.

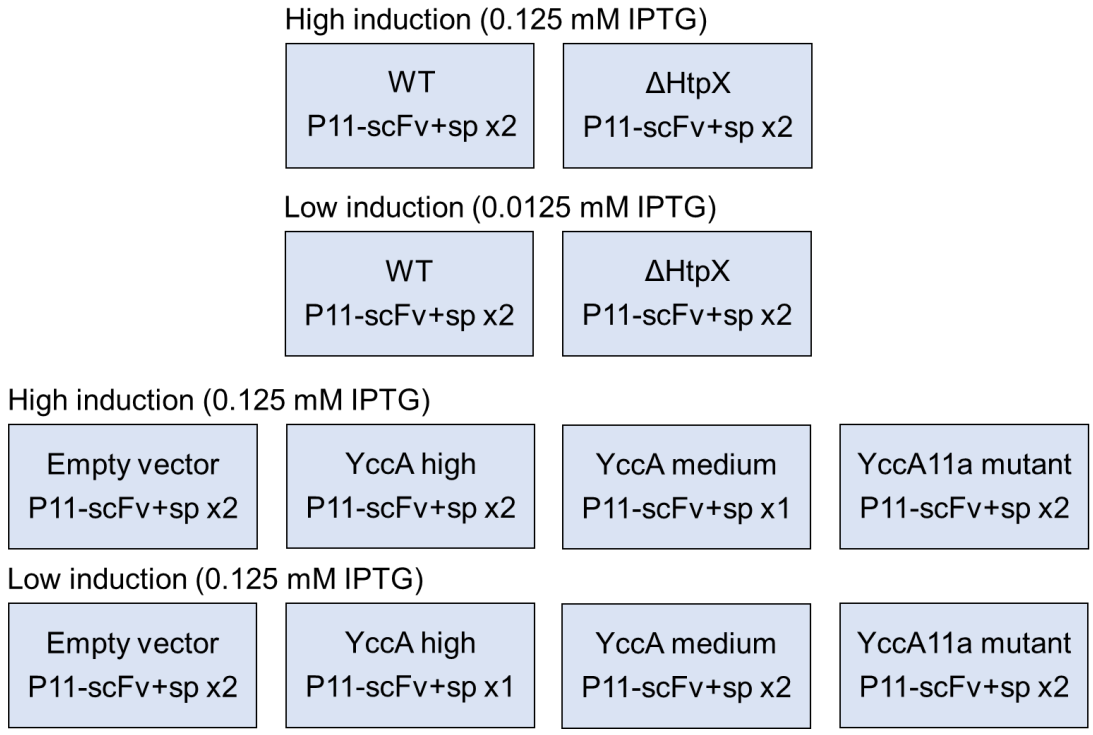


Figure 94. Overview of scFv expression strains tested under Ambr fermenter conditions. The core comparisons are indicated in each row. Samples were induced with high (0.125 mM) or low (0.0125 mM) concentrations of IPTG. Wild-type (WT) and empty vector (pAS15a-Cm) control strains were used as appropriate. The number of biological replicates is also indicated.

4.3.4.2.2 Monitoring growth of engineered strains

Knocking out *htpX* or overexpressing YccA and YccA11 was previously shown to have no impact on growth (see Sections 4.3.3.1.3, 4.3.3.2.2 and 4.3.3.3). However, these studies were conducted in 96 well plate which did not reflect industrially relevant conditions and the scFv was not expressed in these samples. Growth was therefore monitored for all strains tested in the Ambr bioreactors. Biomass production was estimated using OD_{600} (see Appendix 32) and cell weight. The latter provides a more accurate depiction of biomass production, because OD_{600} measurements can be affected by cell size, shape and inclusion body formation.

Growth was maintained for around 24 hours post-induction when inducing high levels of scFv expression (0.125 mM IPTG) before the cultures entered stationary and death phases (Figure 94 A B). In contrast, cultures induced with 0.0125 mM IPTG (lower levels of scFv expression) entered stationary and death phase earlier (around 12 hours post-induction) (Figure 94 C D). This is consistent with previous Ambr fermentations (see section 3.3.2.2) as the secretion system was less overloaded when inducing low levels of scFv secretion and more scFv was present in the periplasm leading to cell leakier/lysing. There was however an increase in growth after this death phase (48 hours post induction). This could be due to the development of mutations that affect scFv expression and/or make *E. coli* that are more resistant to scFv secretion. This hypothesis would however need to be further tested using DNA sequencing.

Knocking out *htpX* had little effect on growth when grown in fermenters when inducing different levels of scFv production (Figure 94 A and C) except 48 hours post induction possibly due to the emergence of resistant strains. Constitutively overexpressing YccA (high and medium) and the YccA11 overactive mutant (medium) has no effect on growth when inducing low amounts of scFv expression compared to the empty vector control, except for 48 hours post induction potentially due to the emergence of resistant strains (Figure 94 D). On the other hand, growth of mutants did vary largely when inducing higher levels of expression of the scFv and when the secretion system

was more overloaded. Overexpressing YccA and YccA11 with a medium strength promoter led to decreased growth rates and a lower maximum density compared to the empty vector control. This could be due to multiple factors, including that these strains grow less well when the secretion system is overloaded and/or that these strains are secreting higher levels of the scFv to the periplasm and are therefore more leaky/lysing.

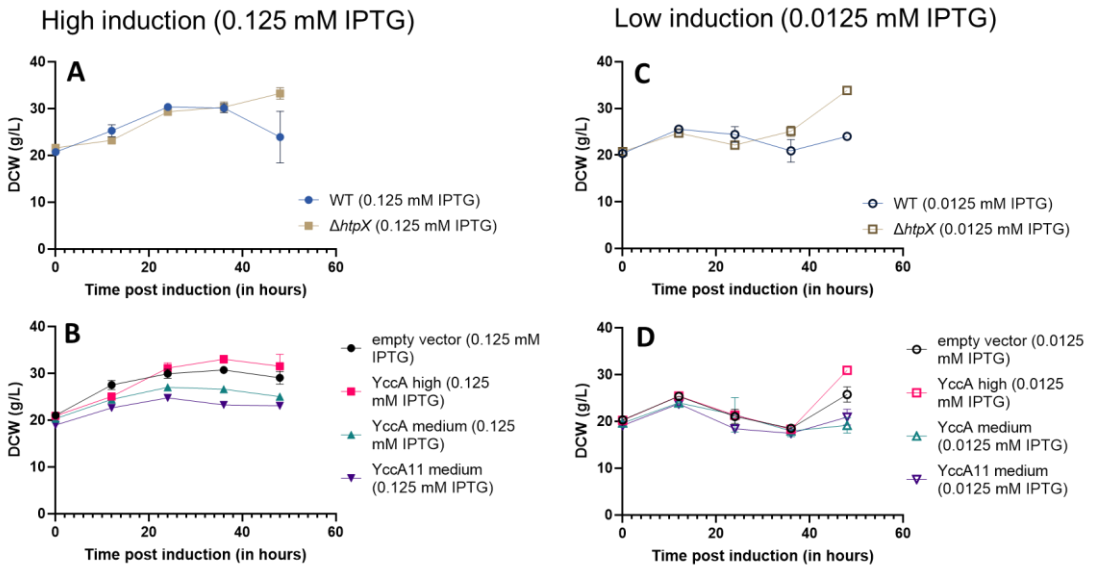


Figure 95. Growth of strains selected for improved scFv production after inducing scFv expression in the periplasm. Expression of the scFv targeted to the periplasm (P11-scFv+sp) was induced with either high (0.125 mM) or low (0.0125 mM) concentrations of IPTG, under Ambr250 high cell density culture conditions. The original parent strain (WT) was used as a control for the isogenic knockout strain (ΔhtpX), and the empty vector (pAS15a-Cm) was the control for strains overexpressing YccA or YccA11 proteins. Biomass production was determined by measuring dry cell weight (DCW). Fermentations were analysed in duplicate except for* YccA *medium (0.125 mM IPTG) and* YccA *high (0.0125 mM IPTG) where only one sample was obtained due to technical issues. Error bars represent one standard deviation.

4.3.4.2.3 Measurements of HCDC scFv production -whole cell analysis

Whole cells were analysed using SDS-PAGE to determine whether mutants improved production and secretion of the scFv under High Cell Density Cultures (HCDC). All SDS-PAGE gels can be found in Appendix 34. For a discussion of the benefits and limitations of SDS-PAGE analysis see section 4.3.4.1.1. The scFv was successfully expressed in all samples as can be seen by the 12 kDa secreted/processed scFv band and 14 kDa unsecreted/unprocessed

scFv band, which were not present in the pre-induced control (Figure 96, Figure 97).

First, the effect of knocking out *htpX* on scFv production and secretion was determined. There was a significant increase in the ratio of secreted (12 kDa) to unsecreted scFv (14 kDa) 24 hours post induction in Δ*htpX* strains induced with 0.125 mM IPTG, based on image analysis (2-way ANOVA, n=2, $p<0.05$) (Figure 96A). However, this difference was not easily observable by eye and no improvements in scFv secretion were observed at all other time points when inducing expression of the scFv with 0.125 mM IPTG which is contrary to observations seen in shake flasks (Figure 96A). Surprisingly inducing low levels of scFv expression (0.0125 mM IPTG), led to a decrease in scFv secretion in *htpX* knockout strains compared to the control, with a significant decrease in the ratio of secreted to unsecreted scFv 12 hours post induction (2-way ANOVA, n=2, $p<0.05$) (Figure 96B). This was contrary to observations in shake flask cultures that showed no difference in secretion of the scFv when induced with low concentrations of IPTG (0.0125 mM) in *htpX* knockout strains. This difference between results in shake flask and HCDC could be due to several factors including differences in cellular responses in different culture conditions, differences in the time points sampled in these two different experiments and the fact that samples were only analysed in duplicate in Ambr bioreactors.

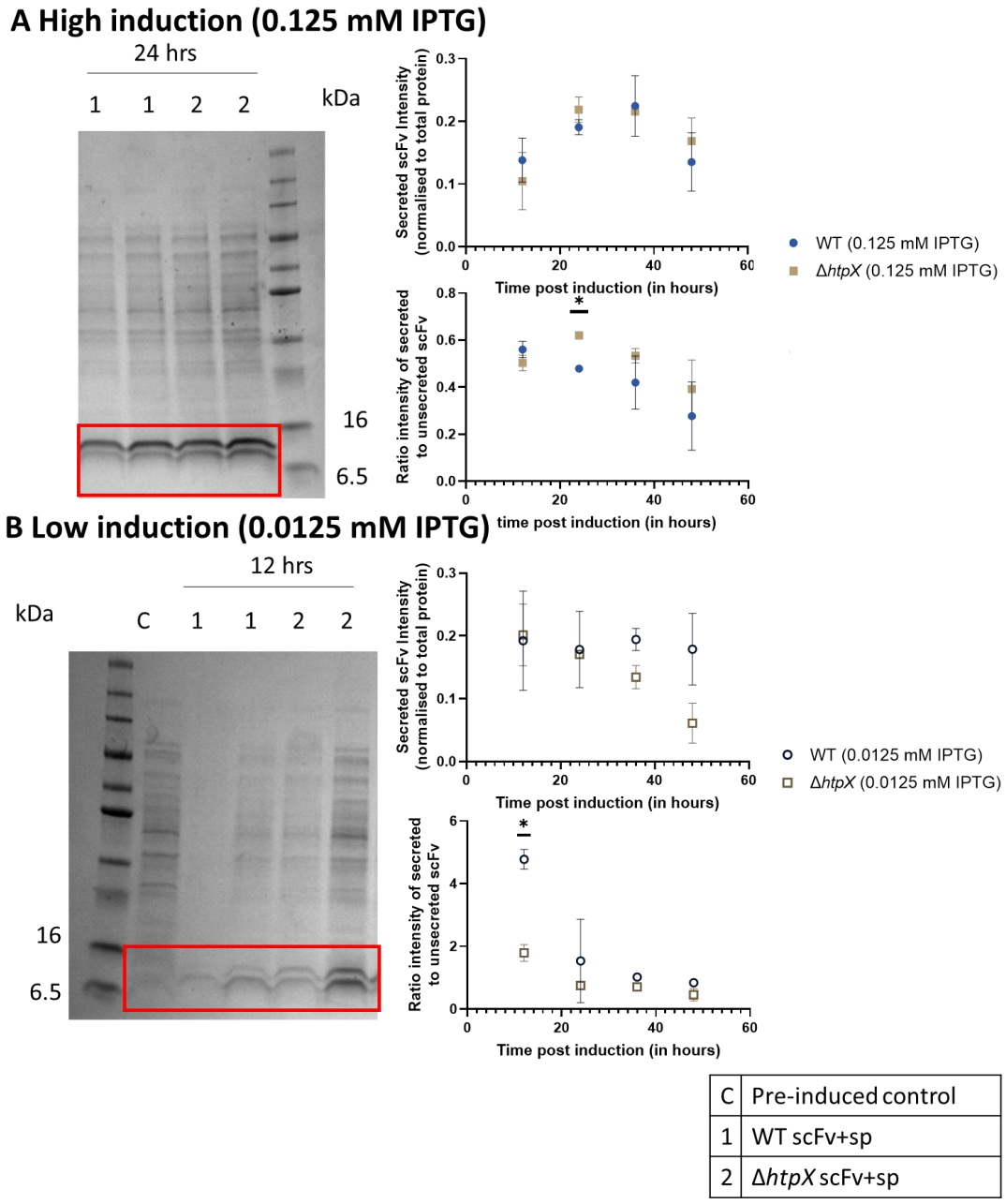


Figure 96.SDS-PAGE of E. coli overexpressing an scFv in the periplasm (scFv+sp) in wild type (WT) and ΔhtpX strains under HCDC conditions induced with high (0.125 mM) (A) and low (0.0125 mM) (B) concentration of IPTG. Sample loading was normalised to bacterial concentration (OD600). The scFv has a predicted molecular weight of 14 kDa with the signal peptide and 12 kDa without the signal peptide and this region of the gels is indicated by the red box. A pre-induced sample was used as a negative control. The ratio of the secreted scFv (12 kDa) to the ratio of unsecreted protein (14 kDa) was determined using densitometry analysis. The 12- and 24-hours and 36- and 48-hours post-induction samples were analysed on separate gels. Statistical analysis was conducted GraphPad Prism using a grouped 2-way ANOVA, with differentially expressed intensities being indicated by a star (N=2). Error bars represent standard deviation.

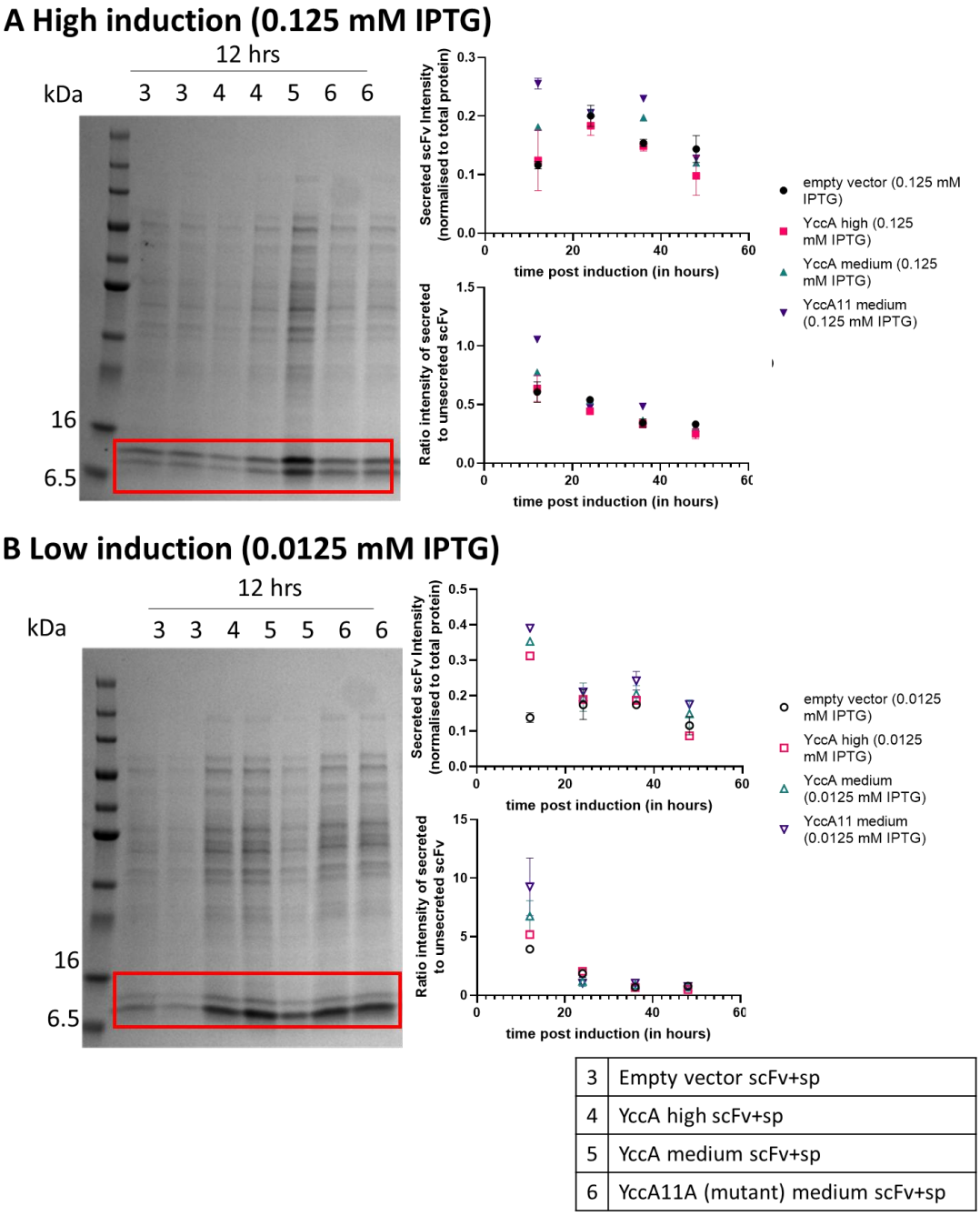
The effect of overexpressing YccA and YccA11 (a more active YccA variant) on scFv expression and secretion was also tested in HCDC fermenter conditions. ScFv expression was induced with two different concentrations of IPTG (0.125 mM) and (0.0125 mM). Unlike shake flask conditions, both high and low expression of the scFv overloaded the secretion system as can be seen by the presence of 14 kDa unsecreted scFv under both conditions. Nonetheless, high levels of scFv expression overloaded the secretion system more as the ratio of unsecreted (14 kDa) to secreted (12 kDa scFv was higher) (Figure 97).

Strains overexpressing YccA11 showed the most promising increase in scFv yields early on in fermentation. Overexpressing YccA11 led to a significant increase in scFv band intensity 24- and 36- hours post induction (0.125 mM IPTG) (2-way ANOVA, n=2, $p<0.05$). The intensity of the secreted scFv (12 kDa) increased 120% and 134% and the ratio of secreted (12 kDa) to unsecreted (14 kDa) scFv increased 74% and 183%, 12 hrs post induction compared to the empty vector when induced with 0.125 and 0.0125 mM IPTG respectively (Figure 97).

An improvement in scFv yields was also seen when overexpressing YccA with medium and high promoters. No improvements were seen in high YccA expression strains when inducing high levels of scFv expression (0.125 mM) however medium YccA overexpression strains led to a maximum 56% increase in the intensity of the secreted scFv bands (12 kDa) and a maximum 29% increase in the ratio of secreted (12 kDa) to unsecreted (14 kDa) scFv (Figure 97A). It was however not possible to statistically determine if YccA medium strains improved scFv yields as only one replicate was available for this sample.

YccA overexpression also improved production of the scFv when induced with lower concentrations of scFv which overloaded the secretion system less (0.0125 mM IPTG). Indeed, there was a 126% and 156% increase in the intensity of the secreted scFv bands (12 kDa) and a 31% and 31% increase in the ratio of secreted (12 kDa) to unsecreted (14 kDa) scFv when overexpressing YccA with high and medium promoters respectively, 12 hrs post induction, which was significantly different when inducing high levels of YccA 12 hours post induction (2-way ANOVA, n=2, $p<0.05$) (0.0125 mM IPTG) (Figure 97B). It was however not possible to determine any statistical difference in scFv abundance when

inducing medium levels of YccA (0.0125 mM IPTG) as only one replicate was available for this sample.



3	Empty vector scFv+sp
4	YccA high scFv+sp
5	YccA medium scFv+sp
6	YccA11A (mutant) medium scFv+sp

Figure 97. SDS-PAGE of E. coli overexpressing an scFv targeted to the periplasm (scFv+sp) in strains constitutively expressing different levels (high and medium) of* YccA *and* YccA11 *at high (0.125 mM) and low (0.0125 mM) concentrations of IPTG. Sample loading was normalised to bacterial concentration (OD600). Gels shown provide an example at a specific time. The scFv has a predicted molecular weight of 14 kDa with the signal peptide and 12 kDa without the signal peptide (this region of the gel is indicated by the red box). A pre-induced sample was used as a negative control. The ratio of the secreted scFv (12 kDa) to the ratio of unsecreted protein (14 kDa) was determined using densitometry analysis. The 12- and 24- hours and 36- and 48-hours post-induction samples were analysed on separate gels. Statistical analysis was conducted using a 2-way ANOVA. Error bars represent standard deviation (N=2).

4.3.4.2.4 Measurement of HCDC scFv secretion – periplasmic extracts

Periplasmic extracts were prepared from the post-induction HCDC samples and analysed using SDS-PAGE to determine whether secretion of the scFv was improved in any of the candidate strains under Ambr conditions. All SDS-PAGE gels can be found in Appendix 35. The 12 kDa band that represents the secreted scFv was identified in all induced cultures (Figure 98, Figure 99); however, there was still a light band with a similar molecular weight in the pre-induced controls which could interfere with scFv quantification. This could be due to the expression plasmid promoter being leaky or a similar molecular weight protein being present in the periplasm.

There was no difference in the intensity of the secreted scFv band (12 kDa) in Δ*htpX* strains compared to the wild-type control (Figure 98). In fact, there was a decrease in the intensity of scFv bands when induced with low concentrations of IPTG (0.0125 mM) which is consistent with SDS-PAGE analysis of whole cells (see section 4.3.4.2.3).

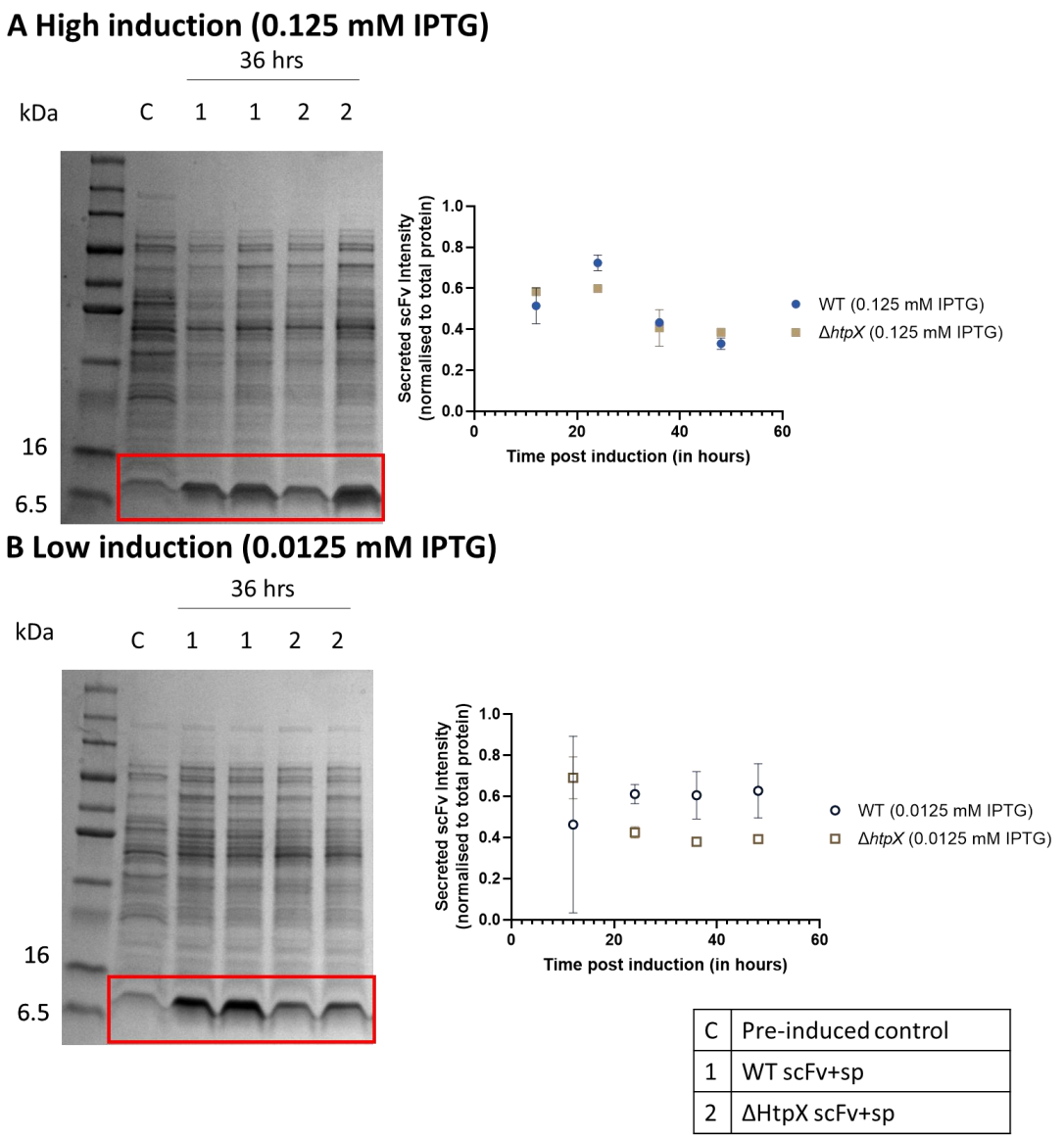
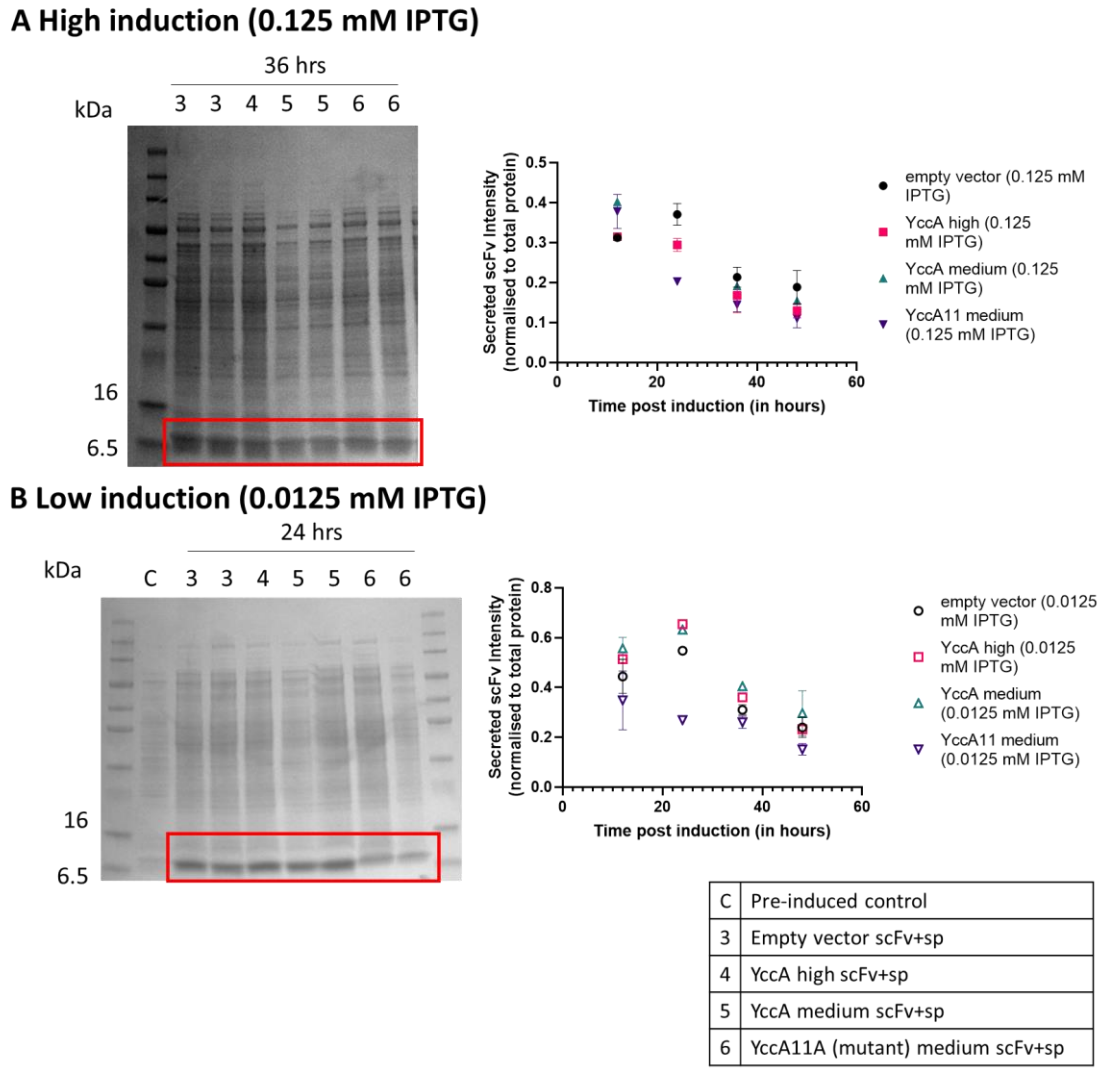


Figure 98. SDS-PAGE analysis of periplasmic extracts obtained from HCDC of ΔhtpX and wild-type strains overexpressing the scFv. Expression of scFv was induced with two concentrations of IPTG (0.125 and 0.0125 mM). A pre-induced sample was also used as a negative control as it should not contain the scFv. Periplasmic extracts were normalised to bacterial concentrations (OD_{600}). The scFv has a predicted molecular weight of 12 kDa without the signal peptide (this region of the gels is indicated by the red box). The band intensity of scFv was quantified using ImageJ and normalised to lane intensity. The 12- and 24- hours and 36- and 48-hours post-induction samples were analysed on separate gels. Error bars represent standard deviation.

No improvements were seen in the quantity of scFv in periplasmic fractions when co-expressing YccA and YccA11 other than a slight increase when inducing expression with 0.0125 mM IPTG and co-expressing high and medium levels of YccA. This contrasts with previous findings from whole cell samples that indicate that more of the scFv is secreted to the periplasm when co-expressing

YccA (see Section 4.3.4.2.3). This could be due to multiple factors. Firstly, increased quantities of scFv in the periplasm is known to increase cell lysis/leakiness, which in turn may mean that less scFv is present in periplasmic fractions and more is present in the culture medium. Secondly, the resolution of the scFv bands is very low in these gels and two bands of a similar molecular weight to the scFv are present in the periplasmic fraction of the pre-induced control, which may be interfering with quantification (Figure 99).



C	Pre-induced control
3	Empty vector scFv+sp
4	YccA high scFv+sp
5	YccA medium scFv+sp
6	YccA11A (mutant) medium scFv+sp

Figure 99. SDS-PAGE analysis of periplasmic extracts obtained from HCDC of strains co-expressing either YccA or a YccA mutant (YccA11) overexpressing the scFv. Expression of scFv was induced with two concentrations of IPTG (0.125 and 0.0125 mM), in duplicate fermentations. A pre-induced sample served as a negative control. The scFv has a predicted molecular weight of 12 kDa without the signal peptide (the relevant region of the gel is boxed in red). The scFv band intensities were quantified using ImageJ and normalised to lane intensity. The 12- and 24- hours and 36- and 48-hours post-induction samples were analysed on separate gels. Error bars represent standard deviation.

4.3.4.2.5 Monitoring cell lysis in engineered strains -supernatant samples

Overexpression of large quantities of proteins targeted to the periplasm often leads to cell leakiness or lysis [42], [91]. Comparison of secreted and unsecreted scFv band intensities in whole cell and periplasmic fractions from HCDC suggested that strains that improved scFv secretion might have higher levels of scFv in the supernatant. Supernatants were therefore analysed using SDS-PAGE to look for the presence of whole cell proteins and the scFv in the supernatant. All SDS-PAGE gels can be found in Appendix 33.

In all strains, leakiness occurred around 24-36 hours after induction with 0.125 mM IPTG, as seen by the presence of whole cell proteins in the supernatant identified using SDS-PAGE (Figure 100, Appendix 33). This was slightly earlier than in our previous Ambr runs but may just be due to variation between fermentation runs (3.3.2.2). On the other hand, inducing expression of the scFv with 0.0125 mM IPTG led to cell lysis/leakiness earlier, as whole cell proteins were found in the supernatant 12-24 hours post induction in all strains (Figure 100, Appendix 33). This is probably due to higher levels of scFv being present in the periplasm. This is also consistent with observations from the growth curves of these samples (Figure 95).

None of the strains were leakier than the control or had higher yields of secreted scFv based on 12 kDa band intensity (Appendix 33), which might be expected if more of the scFv was secreted to the periplasm and could have explained a decrease in growth seen in certain strains when induced with 0.125 mM IPTG (Figure 95). This however may be due to inherent issues when analysing protein abundance using SDS-PAGE.

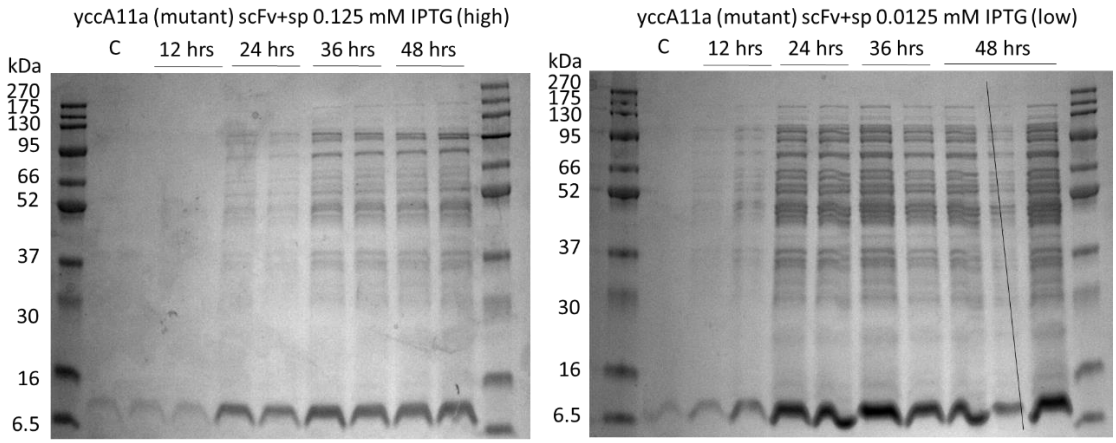


Figure 100. Example SDS-PAGE analyses of supernatants from HCDC co-expressing* YccA11 *from a medium strength constitutive promoter and expressing scFv targeted to the periplasm (scFv+sp). Expression of the scFv was induced with either 0.125 mM IPTG (high) or 0.0125 mM IPTG (low) Samples were taken 12-, 24-, 36- and 48- hours post induction. Bioreactors were run in duplicate. Supernatant samples were normalised to cell density (OD_{600}). A pre-induced sample was used as a negative control (C). One sample was added to the wrong lane and therefore has been crossed out. Samples were analysed on 12% gels.

4.3.4.2.6 Conclusions

Prior work suggested that altering expression of genes involved in SecY and membrane protein degradation improved yields of secreted scFv when induced with 0.125 mM IPTG under shake flask conditions (Δ*htpX*, or overexpressing YccA and YccA11). However, knocking out *htpX* led to no improvements in scFv expression and secretion under HCDC Ambr conditions. However, overexpressing YccA and YccA11 led to improvements a maximum 183% improvement in scFv yields and the ratio of secreted to unsecreted scFv under HCDC Ambr conditions without compromising growth, suggesting that these strains reduce secretion system overloading. This improvement was determined by conducting SDS-PAGE analysis of whole cells, where improvements were particularly prominent in strains overexpressing YccA11 the overactive YccA mutant, early in the process and was seen regardless of the amount of IPTG added (0.125 or 0.0125 mM IPTG). However, surprisingly no increase in scFv abundance in periplasmic fractions or supernatant fractions of strains co-expressing YccA or YccA11. This could be due to several factors including issues with the method of analysis (limited dynamic range etc…), issues with periplasmic extraction methods or due to the low number of

replicates analysed. Nevertheless, growth plateaued earlier in strains overexpressing YccA and YccA11 which is usually an indicator that high levels of proteins are being secreted to the periplasm and cells are lysing.

4.3.4.3 Testing successful strains – production of other recombinant proteins in shake flasks

4.3.4.3.1 Introduction

Based on previous analysis, YccA and YccA11 overexpression strains were the most promising candidates for improving expression and secretion of the scFv in shake flasks and HCDC. However, strains that improve expression of one model protein may not improve expression of other model proteins [99]. These strains were therefore tested for improved expression of multiple disulphide bond containing proteins that are expressed in the periplasm including Hel4 (a long chain antibody fragment), sfGFP (a FLAG tagged non-ATP dependent GFP) and Human Growth hormone (HGH) that were provided by FUJIFILM Diosynth Biotechnologies. All proteins were targeted to the periplasm using an OmpA signal peptide (the same was used for the scFv) and their sequences can be found in Appendix 1. Strains were grown under the shake flask conditions described previously (see methods section 2.2.4) and induced with 0.125 mM IPTG at mid-log phase (OD_{600} 3-4). Hel4 abundance was measured using SDS-PAGE. sfGFP and HGH protein abundance was measured by western blot using anti-FLAG and anti-HGH antibodies, respectively. Production of sfGFP could have been quantified by measuring fluorescence, however, issues identifying the optimal excitation and emission wavelength for this protein meant that this was not used. Nevertheless, images of strains on UV box indicated that the sfGFP was being successfully overexpressed (Appendix 36).

4.3.4.3.2 Testing the effects of YccA and YccA11 co-expression on production of other therapeutic proteins

The effect of YccA overexpression on expression of sfGFP, HGH and Hel4 was tested. YccA overexpression led to a decrease in the intensity of sfGFP and HGH bands compared to the empty vector control in a dose dependent manner

(Figure 101). This could be due to YccA expression negatively affecting the cell and cellular resources used to overexpress proteins, however, further experiments would be needed to test this. There was a significant increase in intensity of normalised Hel4 bands in strains overexpressing YccA, particularly for medium levels of YccA expression (two-way ANOVA, N=6, $p<0.05$) (Figure 101). However, periplasmic extractions would need to be conducted to confirm Hel4 secretion, as only one species was observed in the samples. The YccA overexpression strain is therefore not suitable to improve production of all periplasmic recombinant proteins. Initial data suggests that YccA expression has a negative impact on production of larger proteins (sfGFP, HGH) and not small proteins (<15 kDa) and antibody fragments (scFv, Hel4), however testing of more periplasmic proteins would need to be conducted to confirm this.

The effect of overexpressing YccA11 in a wild-type background was also tested. Western blot analysis suggested an increase in yields of sfGFP and HGH compared to overexpression and empty vector control strains based on band intensity analysis (Figure 102). It was not however possible to determine the fold change increase in these strains as sfGFP and HGH bands were not detected in the parent strains. The effect of overexpressing YccA11 on production of Hel4 was not tested due to time limitations.

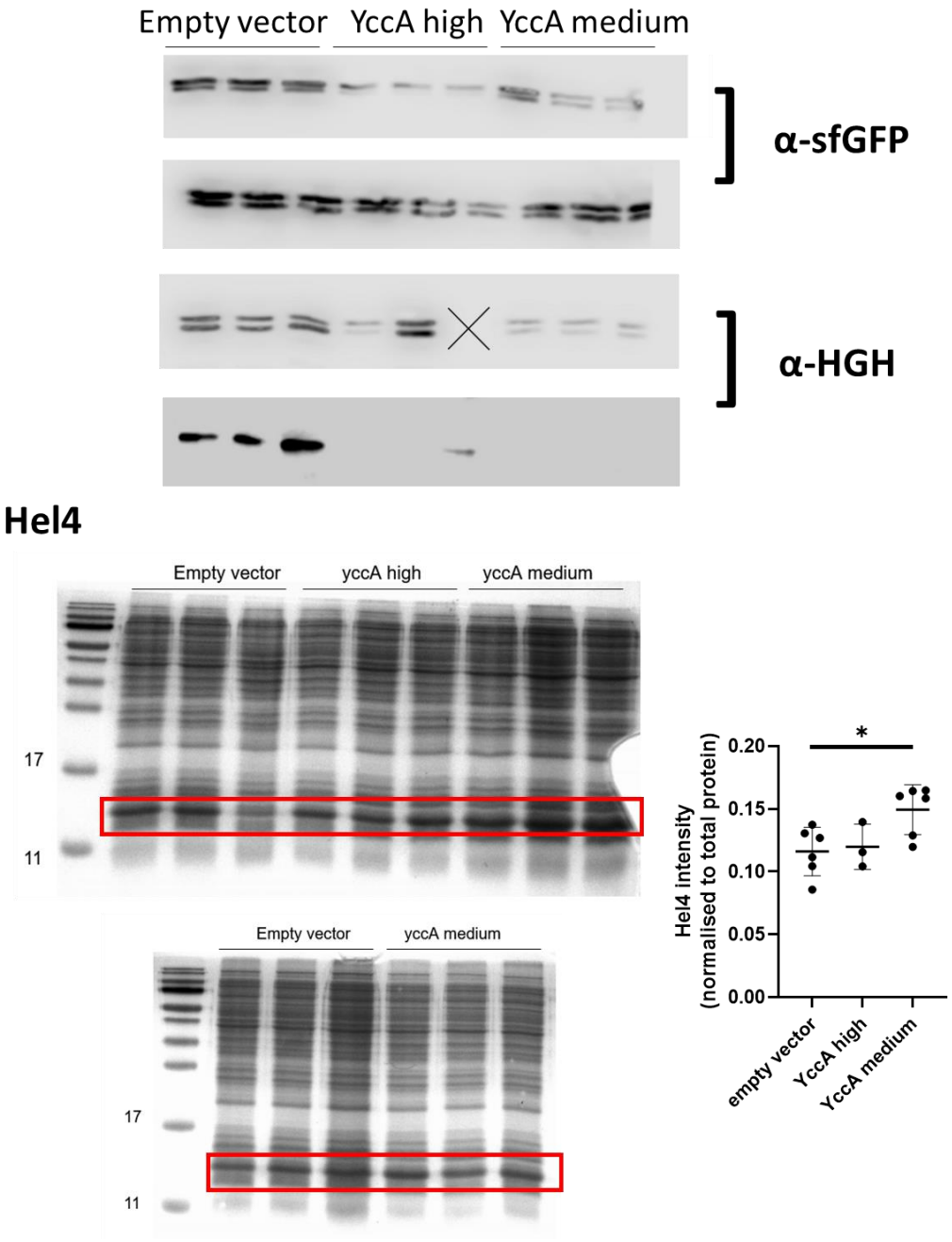


Figure 101. SDS-PAGE and western blot analyses of whole cell samples overexpressing sfGFP. HGH and Hel4 while co-expressing YccA with a high or medium strength promoter compared to an empty vector control (pAS15a_cm). sfGFP was detected using an anti-FLAG tag antibody and is 27-29 kDa depending on whether the signal peptide is present. HGH was detected using an anti HGH antibody and has a molecular weight of 22-24 kDa depending on the presence of the signal peptide. One of the HGH western blots only contains one band which is probably due to insufficient protein separation. Hel4 has a molecular weight of 13-15 kDa and is circled in red. Hel4 abundance was determined using densitometry using ImageJ and normalised to the intensity of each lane and analysed statistically using a one-way ANOVA and statistically different comparisons are indicated by a star. Error bars represent standard deviation. Sample loading was normalised to bacterial concentration (OD_{600}). Samples were tested for equal loading using SDS-PAGE analysis which can be found in Appendix 37.

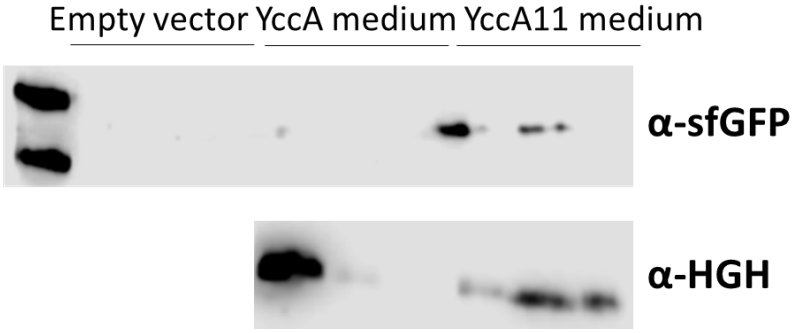


Figure 102. Western blot of whole cell samples overexpressing sfGFP and HGH while co-expressing medium levels of YccA and a YccA overactive mutant (YccA11) compared to an empty vector control (pAS15a_cm). sfGFP was detected using an anti-FLAG tag antibody and is 27-29 kDa depending on whether the signal peptide is present. HGH was detected using an anti HGH antibody and has a molecular weight of 22-24 kDa depending on the presence of the signal peptide. N=3. Sample loading was normalised to bacterial concentration (OD_{600}). Samples were tested for equal loading using a Ponceau stain which can be found in Appendix 37.

4.3.4.3.3 Conclusions

Overexpressing YccA and YccA11 had a positive impact (even if not significantly significant due to low repeat numbers) on expression of a model periplasmic protein (scFv) in shake flask cultures and industrially relevant conditions. Overexpressing YccA11 improved expression of other recombinant periplasmic proteins (sfGFP and HGH); however, expressing YccA had a negative impact on sfGFP and HGH production or no impact on Hel4 production.

4.4 Conclusions- Engineering Strains and medium conditions to improve therapeutic protein production

The OMICs data comparing expression of the scFv proteins targeted to the periplasm or the cytoplasm suggested modifications to the envelope stress response, periplasmic and cytoplasmic chaperones, SecY and membrane protein degradation, amino acid synthesis, flagella synthesis and genes of unknown function. Based on this analysis, strains were created in which 10 individual genes were overexpressed at two different levels, 10 individual genes were deleted, 3 strains were created that combined some of these mutations, and in addition, 3 alterations were made to the growth medium. Most of these engineered strains exhibited no improvements in scFv expression and secretion;

however initial data indicated that strains that inhibited degradation of SecY and membrane protein secretion (Δ*htpX,* and YccA and YccA11 overexpression) improved scFv secretion when the secretion system was overloaded (inducing scFv expression with 0.125 mM IPTG). The Δ*htpX* strains improved scFv expression 90% compared to controls under shake flask conditions when the secretion system was overloaded but did not improve yields of secreted scFv under industrially relevant fermenter conditions. However, YccA and YccA11 overexpression strains were more promising and improved yields of secreted scFv and the ratio of secreted to unsecreted scFv in shake flasks and early on in industrially relevant HCDC (12 hours post induction) improving scFv yields 156% and 134% respectively. Overexpressing YccA improved yields of another periplasmic recombinant protein (Hel4) but had a negative impact on expression of several other proteins (sfGFP and HGH). However, overexpressing YccA11 improved expression of sfGFP and HGH under shake flask cultures. The YccA11 and YccA overexpression strains are therefore the most promising strain generated as part of this study.

5 Chapter 5. Discussion and future prospects.

5.1 Summary of major findings

The work described as part of this study attempts to increase our understanding of how *E. coli* responds to secretion of periplasmic recombinant proteins under industrially relevant conditions and engineer strains and media to improve yields of periplasmic recombinant proteins.

Therapeutic proteins are some of the most expensive treatments on the market compared to small molecule drugs [7] and require lots of trial and error before a commercially viable process is developed [8]. One of the least expensive ways of producing therapeutic proteins is recombinantly in *E. coli* with many therapeutic proteins being secreted to the periplasm to improve yields, ease purification and produce proteins containing disulphide bonds such as antibody fragments which make up >35% of therapeutic proteins produced in *E. coli* [1]. However, overexpressing recombinant proteins in the periplasm often leads to lower yields as the periplasm is limited in size and several issues can occur such as formation of inclusion bodies in the periplasm [66], [93]–[95] and overloading of the secretion system [63], [89], [90], [97], [98]. There is therefore a need to further understand how *E. coli* responds to expression of periplasmic proteins so that novel methods can be developed to improve production of these proteins.

As a prelude to the final discussion, the major findings of each chapter of this thesis are summarised here.

5.1.1 Chapter 3: Multi-OMICs analysis of cellular response to recombinant protein secretion to the periplasm.

In this chapter, a multi-OMICs approach (transcriptomics and proteomics) was used to compare strains expressing a model protein (anti-TNF scFv antibody fragment) in the cytoplasm and periplasm (by adding an OmpA signal peptide), when induced with different concentrations of IPTG (0.125 mM and 0.0125 mM IPTG), at different time points (12 and 24 hours post induction), under industrially relevant High Cell Density Culture (HCDC) conditions. This

increased our knowledge of cellular responses to scFv secretion, secretion system overloading and determine how these changed over time.

- SDS-PAGE analysis of whole cells and periplasmic extracts indicated that inducing scFv expression in the periplasm with 0.125 mM IPTG led to secretion system overloading as more scFv remained unsecreted (14 kDa band) and less scFv was present in periplasmic fractions compared to samples induced with lower levels of IPTG (0.0125 mM).
- Cells expressing the scFv in the periplasm were leakier as indicated by the increase in whole cell proteins in the supernatant identified using SDS-PAGE and the decrease in bacterial growth 24-36 hours post induction that was not present in strains expressing the scFv in the cytoplasm.
- TEM analysis of cells grown in shake flasks showed that insoluble inclusion bodies formed at the poles when the scFv was expressed in the cytoplasm. On the other hand, initial data indicates that inclusion bodies were present near the cytoplasmic inner membrane and in the periplasm when the scFv was secreted. Initial data also indicated that the scFv formed insoluble inclusion bodies when expressed in the periplasm, however these inclusino bodies we
- Transcriptomics and proteomics analysis identified 98% of the transcriptome and 47% of the proteome. Comparing strains expressing the scFv in the cytoplasm and periplasm indicated that scFv secretion affected gene and protein expression particularly when inducing high levels of scFv expression that overloaded the secretion system (0.125 mM IPTG) with 1644 genes and 82 proteins being identified as differentially expressed upon high levels of induction (0.125 mM IPTG) and 592 genes and 30 proteins being identified as differentially expressed upon lower levels of induction (0.0125 mM IPTG). However, few changes in cellular response were seen over time with 70 genes and 281 genes being identified as differentially expressed when comparing strains expressing the scFv in the periplasm 12 and 24 hours post induction when induced with 0.125 mM and 0.0125 mM IPTG respectively.
- Analysis of proteomic and transcriptomic data indicated that scFv secretion led to an increase in several stress responses genes and

proteins (heat shock response, envelope stress response, iron stress, oxidative stress and the SOS response), affected expression of cellular components (overall increased expression of cytoplasmic, membrane and periplasmic protein folding genes, a decreased expression of genes involved in Sec secretion and increased expression flagella genes) and expression of metabolic pathways (decrease in oxidative phosphorylation, decrease in fatty acid synthesis and increase in degradation, amino acid synthesis and degradation and a decrease in ribosome synthesis).

- Most responses described previously were more intense when overloading the secretion system by inducing high levels of scFv secretion (0.125 mM IPTG) however certain responses were unique to secretion system overloading including a decrease in expression of membrane proteins, an increase in expression of pilus genes, a decrease in the CpxR envelope stress response pathway and an increase in expression of osmotic stress genes over time.
- OMICs analysis of strains overexpressing the scFv in the periplasm at different time points (12 hours and 24 hours) indicated that most cellular responses became more intense over time and with strains expressing large quantities of scFv in the periplasm seeing increased expression of osmotic shock genes.

5.1.2 Chapter 4: Engineering strains and media to improve recombinant protein secretion to the periplasm

In this chapter, OMICs data was used to direct engineering of strains and media and was tested to determine if this improved expression of the scFv under shake flask conditions that were optimised to reproduce industrial conditions as much as possible. Successful strains were then tested to see if they improved expression of the scFv under industrially relevant fermenter conditions and see if they improved expression of other periplasmic recombinant proteins (HGH, Hel4 and sfGFP).

- A scale down model was developed in shake flasks to test strain and media alterations in a high throughput manner. This was developed by using culture medium that was similar to Ambr fermentations by reducing culture volumes to 20 mL to improve aeration as this was shown to improve the final bacterial concentration (OD_{600}) 0.6-fold (but not adding carbon, nitrogen or adjusting the pH). Under these conditions srains were also leaky when overexpressing the scFv in the periplasm as can be seen by the presence of the secreted scFv in the supernatant 3 hours post induction.
- Based on OMICs data, the envelope stress response pathway, periplasmic and cytoplasmic chaperones, SecY and membrane protein proteases, amino acids synthesis pathways, flagella synthesis pathways, and genes of unknown function were altered by overexpressing 10 genes at 2 different levels, knocking out 10 genes, creating 3 strains that combined these mutations and making 3 alterations to media composition which were shown to have no effect on growth.
- Most of these strains did not improve expression and secretion of the scFv based on SDS-PAGE analysis of whole cells when induced with 0.125 mM IPTG and grown under shake flask conditions other than those that decreased expression and activity of proteins involved in SecY and membrane protein degradation (Δ*htpX,* YccA and YccA11 overexpression).
- Image analysis of SDS-PAGE analysis indicated that knocking out the membrane protease HtpX improved scFv expression 90% compared to controls under shake flask conditions when the secretion system was overloaded (0.125 mM IPTG) but did not improve yields of secreted scFv under industrially relevant fermenter conditions or when the secretion system was less overloaded (induced with 0.0125 mM IPTG).
- SDS-PAGE analysis of whole cells indicated that overexpressing negative regulators of membrane protein degradation YccA and the more active mutant YccA11, improved yields of secreted scFv 77% and 50-70% respectively in shake flasks cultures but only when inducing high levels of scFv secretion that overloaded the Sec secretion system (0.125 mM IPTG).

- Overexpressing YccA and YccA11 also improved intensity of scFv bands based on image analysis of whole cell proteins 156% and 134% respectively under industrially relevant high cell density cultures. However, this improvement was only seen 12 hours into the fermentation run and did not improve final scFv yields at the end of the process. Furthermore, improvements were only identified based on SDS-PAGE analysis of whole cells and not periplasmic fractions.
- Overexpressing YccA potentially improved expression of Hel4 but had a negative impact on expression of several other recombinant proteins in the periplasm (sfGFP and HGH) based on western blot analysis under shake flask conditions based on Western Blot and SDS-PAGE analysis. However overexpressing YccA11 improved expression of several other proteins (sfGFP and HGH) under shake flask cultures.

5.2 OMICs: a successful tool to study cellular responses to recombinant protein secretion.

This thesis also takes a novel approach by using multi-OMICs (transcriptomics and proteomics) to study cellular responses when overexpressing periplasmic recombinant proteins which provides a more comprehensive view of cellular physiology. Furthermore, to our knowledge is the first OMICs study to compare *E. coli* strains expressing the same recombinant protein in different cellular compartments which allowed us to determine cellular responses that are specifically associated with recombinant protein secretion to the periplasm and secretion system overloading and are not due to protein overexpression. This scFv was also insoluble in the cytoplasm and predominantly soluble in the periplasm as indicated by SDS-PAGE analysis of soluble and insoluble fractions. TEM analysis of samples overexpressing the scFv in the cytoplasm and periplasm under shake flask conditions indicated that some inclusion bodies may be forming in the periplasm when secreting the scFv. However, these inclusion bodies are difficult to observe as they are a lot smaller than those found when expressing large quantities of insoluble proteins in the periplasm [93], [95], [258]. Ideally, TEM would have been conducted on samples from Ambr fermenters and later on in the fermentation run where more of the

processed secreted scFv was present in the insoluble fraction as these inclusion bodies would have potentially been larger and easier to observe however this was not possible as these samples had been frozen.

There are however some important factors to consider when interpreting this OMICs data. Firstly, most differentially expressed genes identified in this study were identified using transcriptomics, however transcript abundance does not necessarily correlate with protein abundance [319]. This is in part due to the fact that only a portion of the proteome can be analysed using proteomics [177], [263]. It is also important to consider that iTRAQ proteomics underestimates fold change information due to precursor interference [184], [187]–[189] unless using specialised equipment that was not available for this study [191], [192]. Nonetheless, iTRAQ provides excellent information about whether a protein is upregulated or downregulated and requires shorter mass spectrometry run times hence why it was used as part of this study.

OMICs analysis also presents several challenges which are important to consider when interpreting OMICs data. Firstly databases used for enrichment analysis such as Gene Ontology [255] and KEGG are often poorly annotated and use classifications that are highly biased [264]. This study therefore also used *E. coli* specific tools such as EcoCyc [251], [252] and Tfinfer [254], however, issues still occurred (e.g. TFinfer analysis did not identify the PspF regulon as being upregulated upon scFv secretion). To overcome this data was also supplemented with information from other databases such as RegulonDB [266] and STEPdb [265] however this curation was time consuming. Secondly, OMICs analysis is limited by our current knowledge of gene function which is limited as around 1/3rd of genes have no known function [288], genes often have multiple functions and our knowledge of gene regulation is highly limited. Thirdly, integrating multiple OMICs datasets is challenging. The advantages and disadvantages of each OMICs method used were therefore considered [263] and tools were used to map multiple different types of OMICs data [275].

Furthermore, proteomics only quantifies protein abundance and not necessarily protein activity which can be influenced by protein cellular localisation, protein interactions, post translational modifications and metabolite abundance. In future work, it would therefore be interesting to determine the

effect of scFv secretion on protein function by using methods such as linking mass spectrometry, spatial proteomics, enrichment of post-translational modifications, metabolomics, and metabolic modelling.

5.3 Drawing broader conclusions from OMICs data.

While this OMICs data has provided useful information about how *E. coli* responds to secretion of a model protein (an anti-TNF scFv antibody fragment) under specific culture conditions it is important not to draw broader conclusions about cellular responses to secretion of all periplasmic recombinant proteins, in all strains and under all culture conditions.

Indeed, cellular responses have been shown to vary when overexpressing different recombinant proteins in *E. coli* [44], [107], [124], [148], [168], [200], [203], [205], [211], [221], [230]. This also means that the addition of the C-terminal OmpA signal peptide to the secreted form of the scFv may also be a contributing factor to changes in cellular responses, however this signal peptide is small (63 bp) and is cleaved off in the periplasm. Cellular responses are also known to vary based on the strain used and culture conditions [21], [169], [170], [204], [219], [223], [225], [232], [234]. This variation is however an inherent issue with OMICs studies. It would have been interesting to observe how *E. coli* responds to expression of periplasmic recombinant proteins under different conditions, however, this was not the main aim of this study.

Because of this variation in cellular responses, multiple steps were taken when designing this study to ensure that OMICs data was as applicable as possible to real world bioprocesses. This included using an industrially relevant therapeutic protein (an anti-TNF scFv antibody) and using the same strain and culture conditions as those used at FDB (W3110 based strain, rich media, High Cell Density Cultures).

5.4 Testing engineered strains and media was challenging

OMICs data was then used to create novel strains and media modifications. Alterations to media and strains can be broadly grouped into multiple categories: altering expression of the envelope stress response pathway, decreasing expression and activity of proteins involved in SecY and

membrane protein degradation, decreasing expression of flagella genes, adding amino acids to media and altering expression of genes of unknown function. Specific targets were identified based on whether a gene was differentially expressed at the transcript and protein level, whether similar responses were seen in the different conditions tested, their fold change level and information from the literature. In retrospect, a more systematic approach could have been taken to select targets such as the Multi-OMIC based production strain improvement (MOBpsi) method that classifies potential targets using decision trees [263].

The number of media and strains tested in this study was considerably more than reported in previous studies that combined OMICs and genetic engineering to improve recombinant protein overexpression [146], [201], [207], [236]. However, most of the alterations tested did not improve expression of the scFv under shake flask conditions.

Indeed, there were limitations in the methods used to test strains and medium alterations. Firstly, culture conditions used to test strains (shake flask cultures) were different to those used to generate OMICs data (fed batch Ambr fermenters) which can affect cellular physiology. Multiple efforts were taken to reproduce culture conditions used in fed batch fermenters in shake flasks by using similar media and increasing bacterial density by reducing culture volumes and improving aeration. In future work, it would therefore be interesting to use OMICs to determine the similarities and differences in cellular responses to scFv overexpression and secretion under these two different culture conditions. Interestingly cells expressing the scFv in the periplasm were leaky when grown in shake flask cultures. Cell leakiness can aid in purification as proteins can be extracted from the supernatant which requires less sample processing and reduces sample complexity but has been predominantly reported in industrial fermentation processes [42]. This increased leakiness may be due to several factors including high levels of scFv expression in the periplasm, however this may be useful when purifying recombinant proteins on a smaller scale. Secondly it was not possible to confirm that proteins were successfully overexpressed in the correct cellular compartment in all overexpression strains as antibodies were not easily available for these proteins other than PspA.

Therefore, overexpression strains may simply not be overexpressing our genes of interest.

Furthermore, due to time limits strains were only tested when induced with high levels of IPTG (0.125 mM IPTG) that overloaded the secretion system. In future work it would therefore be interesting to test engineered strains and media when induced with lower levels of IPTG that overloaded the secretion system less. Finally, improvements in scFv secretion were primarily quantified by analysing whole cell proteins due to time limitations. In future work, it would therefore be useful to test periplasmic fractions and supernatant fractions. Furthermore, controlled leakiness is a characteristic that is sought after when expressing periplasmic recombinant proteins and it would therefore be interesting to test strains to determine if they are leaky using fluorescent or colorimetric assays [42].

It is also important to consider the limitations of quantifying scFv abundance using SDS-PAGE analysis. SDS-PAGE can have issues with dynamic range, background, unequal loading, and band resolution which can make it difficult to observe small changes. Image analysis of normalised scFv band intensities were used to overcome some of these issues however this still presents several challenges [316], [317]. Antibody based methods (western blot, ELISA etc…), activity assays or quantification of purified scFv would have provided a more accurate quantifications of scFv abundance and would have improved our ability to observe smaller improvements in scFv expression and secretion, however, antibodies were not available for the scFv and when the scFv was tagged with a his-tag this significantly affected recombinant protein overexpression. None the less SDS-PAGE is commonly used in industry to evaluate different strains/media conditions as it is rapid, requires low volumes of samples, is applicable to all recombinant proteins, is cheap and allows for the detection of large changes in recombinant protein abundance hence why it was used in this study.

5.5 Scope to further engineer strains

While testing of engineered strains and media was challenging, several modifications (Δ*htpX,* and YccA and YccA11 overexpression) were developed

that potentially improved expression of the scFv under shake flask conditions when the secretion system was overloaded (scFv expression induced with 0.125 mM IPTG). This is consistent with a previous study that identified that overexpressing YccA and YccA11 improved expression and secretion of LamB-LacZ and lowered secretion system overloading [280]. This is however to our knowledge the first time that these strains have been identified as improving expression of recombinant periplasmic proteins. In future work complementation of these mutants could be conducted to further verify improvements seen. Furthermore, it would also be interesting to further analyse these mutants by conducting OMICs analysis to identify how cellular responses change in these strains and use SecY western blots to verify if these strains inhibit SecY degradation as predicted.

These successful strains were also tested to determine if they improved expression of recombinant proteins under industrially relevant conditions. However, this step is often omitted from strain engineering studies. Overexpression of YccA and YccA11 improved expression of the scFv industrial processes within the first 12 hours post induction, however, *htpX* knockouts did not. This is expected as methods developed to improve recombinant protein overexpression do not always translate to other bioprocesses. Furthermore, overexpression of YccA11 improved expression of HGH and sfGFP and initial data indicated that YccA overexpression may slightly improve Hel4 yields. This is also expected as methods that improve expression of one recombinant protein often improve expression of others [99]. Indeed modifications that had previously been identified as improving expression of periplasmic recombinant proteins did not improve expression of the scFv in this study which could be due to different proteins being overexpressed and culture conditions [131], [146]–[148].

In future work, it would be interesting to characterise why YccA and YccA11 overexpression did not improve expression of other recombinant periplasmic proteins. It would be interesting to further characterise processes where sfGFP, HGH and Hel4 are overexpressed to determine if the secretion system was overloaded and conduct OMICs analysis of strains overexpressing

these proteins to determine if there is also an increase in expression of genes involved in SecY and membrane protein degradation.

Overall overexpressing YccA and the YccA11 mutant was identified as the most promising strategy to improve recombinant protein secretion, however, improvements in expression of recombinant periplasmic expression were not as high as expected (55-156% increase in yields of secreted scFv) and not sufficient to use this strain in industrial bioprocesses. Indeed other studies using OMICs data to direct strain engineering saw 63% [146], 125% [201] or even 480% [207] improvements in expression of specific recombinant proteins. There is however scope to further improve periplasmic protein expression. It would be interesting to alter the wildtype *yccA* using mutagenesis or to overexpress YccA11 in a Δ*yccA* background. Indeed, initial data indicated that overexpressing YccA11 in a Δ*yccA* background improved scFv expression in shake flasks compared to a wildtype background, however further tests and replicates would need to be conducted to confirm this. Secondly, YccA11 was only overexpressed with the medium level Anderson promoter, therefore higher and lower levels of YccA11 could be tested to determine if this further improves recombinant protein overexpression. The YccA and YccA11 overexpression strains could also be combined with other mutations such as those that also inhibit SecY and membrane protein degradation (HflCK overexpression and knocking out *htpX*) or combined with SecY overexpression. Furthermore, OMICs analysis of strains that improved expression of periplasmic recombinant proteins could be used to direct strain engineering.

5.6 Other potential applications of OMICs data

Furthermore, there are other potential applications of OMICs data that could be tested in future work. Alterations to strains and medium could be tested such as increasing expression of CydDC a component of the electron transport chain involved in cysteine transport and redox balance, increasing expression of the periplasmic chaperones IbpAB which were shown to be highly upregulated at the proteomic and transcriptomic level upon scFv secretion, increasing expression of periplasmic protein folding components that were affected by scFv secretion including DsbAB and FkpA, increasing expression of Sec secretion

system components (SecYEG), altering expression of genes of unknown function (*yaiY, ybhQ, yggE* and *yciF*) that are either related to the envelope stress response pathway [78], [291] or have been identified as being differentially expressed in other OMICs studies looking at expression of periplasmic recombinant proteins [273], [286].

OMICs data could also be used to generate negative feedback loops [320] or develop reporters [200], [321]. In particular, methods could be developed to detect and regulate expression of recombinant proteins based on secretion system overloading would be of particular interest. Expression of pilus proteins and the CpxR response would be a potentially interesting target for this as changes in these stress responses were only seen when inducing high levels of scFv expression that overloaded the secretion system (0.125 mM IPTG).

5.7 Conclusions

By comparing expression of a model protein (scFv) in the cytoplasm and periplasm this study has increased our understanding of cellular responses that are specific to secretion and secretion system overloading under industrially relevant conditions. This data has been successful applied to generate and test a wide range of novel methods to improve expression of recombinant proteins in the periplasm including altering expression of genes involved in SecY and membrane protein degradation (Δ*htpX,* and YccA and YccA11 overexpression) which showed improved scFv expression and secretion under shake flask conditions. The YccA and YccA11 overexpression strains were the most promising as initial data indicated that this strain improved yields of the secreted scFv 156 and 134% respectively under industrially relevant conditions and improved expression of other periplasmic recombinant proteins (Hel4, HGH and sfGFP). However, YccA11 overexpression strains only improved scFv expression early on in industrial processes and did not improve final yields at the end of the batch. Nonetheless, there is still scope to further improve YccA11 overexpression strains and use OMICs data generated in this study to further develop methods to improve expression of periplasmic recombinant proteins. Improving expression of periplasmic recombinant proteins has the potential to reduce the amount of trial and error and cost associated with developing a

process for production of therapeutic proteins that are secreted to the periplasm such as antibody fragments.

6 References

[1] T. S. Castiñeiras, S. G. Williams, A. G. Hitchcock, and D. C. Smith, "*E. coli* strain engineering for the production of advanced biopharmaceutical products," *FEMS Microbiology Letters*. 2018, doi: 10.1093/femsle/fny162.

[2] K. Itakura *et al.*, "Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin," *Science (80-.).*, 1977, doi: 10.1126/science.412251.

[3] D. V Goeddel *et al.*, "Expression in *Escherichia coli* of chemically synthesized genes for human insulin.," *Proc. Natl. Acad. Sci. U. S. A.*, 1979.

[4] D. V. Goeddel *et al.*, "Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone," *Nature*, 1979, doi: 10.1038/281544a0.

[5] G. Walsh, "Biopharmaceutical benchmarks 18," *Nat. Biotechnol.*, 2018, doi: 10.1038/nbt0706-769.

[6] "Recombinant Protein Market - Growth, Trends, Covid-19 Impact, and Forecasts (2021 - 2026) - GII." https://www.giiresearch.com/report/moi707231-global-recombinant-protein-market-segmented-by.html (accessed Jul. 27, 2022).

[7] F. D. Makurvet, "Biologics vs. small molecules: Drug costs and patient access," *Med. Drug Discov.*, vol. 9, p. 100075, Mar. 2021, doi: 10.1016/J.MEDIDD.2020.100075.

[8] G. L. Rosano and E. A. Ceccarelli, "Recombinant protein expression in *Escherichia coli*: advances and challenges," *Front. Microbiol.*, vol. 5, 2014, doi: 10.3389/fmicb.2014.00172.

[9] P. Savitsky *et al.*, "High-throughput production of human proteins for crystallization: the SGC experience," *J. Struct. Biol.*, vol. 172, no. 1, pp. 3–13, Oct. 2010, doi: 10.1016/J.JSB.2010.06.008.

[10] D. Esposito and D. K. Chatterjee, "Enhancement of soluble protein expression through the use of fusion tags," *Curr. Opin. Biotechnol.*, vol. 17, no. 4, pp. 353–358, Aug. 2006, doi: 10.1016/J.COPBIO.2006.06.003.

[11] S. Y. Lee, "High cell-density culture of *Escherichia coli*," *Trends in Biotechnology*. 1996, doi: 10.1016/0167-7799(96)80930-9.

[12] L. Yee and H. W. Blanch, "Recombinant protein expression in High Cell-Density fed-batch cultures of *Escherichia coli*.," *Bio-Technology*, vol. 10, no. 12, pp. 1550–1556, 1992, doi: 10.1038/nbt1292-1550.

[13] J. Shiloach and R. Fass, "Growing *E. coli* to high cell density - A historical perspective on method development," *Biotechnology Advances*. 2005, doi: 10.1016/j.biotechadv.2005.04.004.

[14] J. H. Choi, K. C. Keum, and S. Y. Lee, "Production of recombinant proteins by high cell density culture of *Escherichia coli*," *Chem. Eng. Sci.*, vol. 61, no. 3, pp. 876–885, 2006, doi: 10.1016/j.ces.2005.03.031.

[15] T. Selas Castiñeiras, S. G. Williams, A. Hitchcock, J. A. Cole, D. C. Smith, and T. W. Overton, "Development of a generic β-lactamase screening system for improved signal peptides for periplasmic targeting of recombinant proteins in *Escherichia coli*," *Sci. Reports 2018 81*, vol. 8, no. 1, pp. 1–18,

May 2018, doi: 10.1038/s41598-018-25192-3.

[16] N. Ferrer-Miralles, J. Domingo-Espín, J. Corchero, E. Vázquez, and A. Villaverde, "Microbial factories for recombinant pharmaceuticals," *Microbial Cell Factories*, vol. 8. p. 17, Mar. 24, 2009, doi: 10.1186/1475-2859-8-17.

[17] K. Hayashi *et al.*, "Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110," *Mol. Syst. Biol.*, vol. 2, 2006, doi: 10.1038/msb4100049.

[18] B. Ou, C. Garcia, Y. Wang, W. Zhang, and G. Zhu, "Techniques for chromosomal integration and expression optimization in *Escherichia coli*," *Biotechnol. Bioeng.*, 2018, doi: 10.1002/bit.26790.

[19] U. Mamat *et al.*, "Endotoxin-free protein production—ClearColi™ technology," *Nat. Methods 2013 109*, vol. 10, no. 9, pp. 916–916, Aug. 2013, doi: 10.1038/nmeth.f.367.

[20] U. Mamat *et al.*, "Detoxifying *Escherichia coli* for endotoxin-free production of recombinant proteins," *Microb. Cell Fact.*, vol. 14, no. 1, pp. 1–15, Apr. 2015, doi: 10.1186/S12934-015-0241-5/FIGURES/10.

[21] J. Pandhal, S. Y. Ow, J. Noirel, and P. C. Wright, "Improving N-Glycosylation Efficiency in *Escherichia coli* Using Shotgun Proteomics, Metabolic Network Analysis, and Selective Reaction Monitoring," *Biotechnol. Bioeng.*, vol. 108, no. 4, pp. 902–912, 2011, doi: 10.1002/bit.23011.

[22] A. M. Sanden *et al.*, "Limiting factors in *Escherichia coli* fed-batch production of recombinant proteins," *Biotechnol. Bioeng.*, vol. 81, no. 2, pp. 158–166, 2003, doi: 10.1002/bit.10457.

[23] N. Sethuraman and T. A. Stadheim, "Challenges in therapeutic glycoprotein production," *Current Opinion in Biotechnology*. 2006, doi: 10.1016/j.copbio.2006.06.010.

[24] S. R. P. Jaffé, B. Strutton, Z. Levarski, J. Pandhal, and P. C. Wright, "*Escherichia coli* as a glycoprotein production host: recent developments and challenges," *Curr. Opin. Biotechnol.*, vol. 30, pp. 205–210, Dec. 2014, doi: 10.1016/J.COPBIO.2014.07.006.

[25] S. - European Unions Horizon *et al.*, "Microbial protein cell factories fight back?," *Trends Biotechnol.*, vol. 0, no. 0, Dec. 2021, doi: 10.1016/J.TIBTECH.2021.10.003.

[26] A. Sandomenico, J. P. Sivaccumar, and M. Ruvo, "Evolution of *Escherichia coli* Expression System in Producing Antibody Recombinant Fragments," *Int. J. Mol. Sci.*, vol. 21, no. 17, pp. 1–39, Sep. 2020, doi: 10.3390/IJMS21176324.

[27] A. Mullard, "FDA approves 100th monoclonal antibody product," *Nat. Rev. Drug Discov.*, vol. 20, no. 7, pp. 491–495, Jul. 2021, doi: 10.1038/D41573-021-00079-7.

[28] S. Jin *et al.*, "Emerging new therapeutic antibody derivatives for cancer treatment," *Signal Transduct. Target. Ther. 2022 71*, vol. 7, no. 1, pp. 1–28, Feb. 2022, doi: 10.1038/s41392-021-00868-x.

[29] G. Rodrigo, M. Gruvegård, and J. M. Van Alstine, "Antibody Fragments and Their Purification by Protein L Affinity Chromatography," *Antibodies 2015, Vol. 4, Pages 259-277*, vol. 4, no. 3, pp. 259–277, Sep. 2015, doi:

10.3390/ANTIB4030259.

[30] H. Nakamoto and J. C. A. Bardwell, “Catalysis of disulfide bond formation and isomerization in the *Escherichia coli* periplasm,” *Biochimica et Biophysica Acta - Molecular Cell Research*. 2004, doi: 10.1016/j.bbamcr.2004.02.012.

[31] H. Kadokura, F. Katzen, and J. Beckwith, “Protein disulfide bond formation in prokaryotes,” *Annu. Rev. Biochem.*, vol. 72, pp. 111–135, 2003, doi: 10.1146/annurev.biochem.72.121801.161459.

[32] F. Hatahet, V. D. Nguyen, K. E. H. Salo, and L. W. Ruddock, “Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli*.,” *Microb. Cell Fact.*, 2010, doi: 10.1186/1475-2859-9-67.

[33] C. F. R. O. Matos *et al.*, “Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains,” *Biotechnol. Prog.*, 2014, doi: 10.1002/btpr.1858.

[34] A. Gaciarz *et al.*, “Systematic screening of soluble expression of antibody fragments in the cytoplasm of *E. coli*,” *Microb. Cell Fact.*, vol. 15, no. 1, Jan. 2016, doi: 10.1186/S12934-016-0419-5.

[35] A. Gaciarz *et al.*, “Efficient soluble expression of disulfide bonded proteins in the cytoplasm of *Escherichia coli* in fed-batch fermentations on chemically defined minimal media,” *Microb. Cell Fact.*, vol. 16, no. 1, pp. 1–12, Jun. 2017, doi: 10.1186/S12934-017-0721-X/TABLES/2.

[36] J. Lobstein, C. A. Emrich, C. Jeans, M. Faulkner, P. Riggs, and M. Berkmen, “SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm,” *Microb. Cell Fact.*, vol. 11, no. 1, pp. 1–16, May 2012, doi: 10.1186/1475-2859-11-56/TABLES/3.

[37] P. H. Bessette, F. Åslund, J. Beckwith, and G. Georgiou, “Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 24, pp. 13703–13708, Nov. 1999, doi: 10.1073/PNAS.96.24.13703/ASSET/E2247143-A591-4CD7-A762-9DD2C35509CF/ASSETS/GRAPHIC/PQ2494137005.JPEG.

[38] P. Cornelis, “Expressing genes in different *Escherichia coli* compartments,” *Current Opinion in Biotechnology*. 2000, doi: 10.1016/S0958-1669(00)00131-2.

[39] F. J. M. Mergulhão, D. K. Summers, and G. A. Monteiro, “Recombinant protein secretion in *Escherichia coli*,” *Biotechnology Advances*. 2005, doi: 10.1016/j.biotechadv.2004.11.003.

[40] T. Loo, M. L. Patchett, G. E. Norris, and J. S. Lott, “Using Secretion to Solve a Solubility Problem: High-Yield Expression in *Escherichia coli* and Purification of the Bacterial Glycoamidase PNGase F,” *Protein Expr. Purif.*, vol. 24, no. 1, pp. 90–98, Feb. 2002, doi: 10.1006/PREP.2001.1555.

[41] H. C. Neu and L. A. Heppel, “The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts.,” *J. Biol. Chem.*, 1965.

[42] J. Kastenhofer, V. Rajamanickam, J. Libiseller-Egger, and O. Spadiut, “Monitoring and control of *E. coli* cell integrity,” *J. Biotechnol.*, vol. 329, pp.

1–12, Mar. 2021, doi: 10.1016/J.JBIOTEC.2021.01.009.

[43] J. Oswald, R. Njenga, A. Natriashvili, P. Sarmah, and H. G. Koch, "The Dynamic SecYEG Translocon," *Front. Mol. Biosci.*, vol. 8, Apr. 2021, doi: 10.3389/FMOLB.2021.664241.

[44] I. Guerrero Montero *et al.*, "Comparative proteome analysis in an *Escherichia coli* CyDisCo strain identifies stress responses related to protein production, oxidative stress and accumulation of misfolded protein," *Microb. Cell Fact.*, 2019, doi: 10.1186/s12934-019-1071-7.

[45] P. Natale, T. Brüser, and A. J. M. Driessen, "Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms," *Biochimica et Biophysica Acta - Biomembranes*. 2008, doi: 10.1016/j.bbamem.2007.07.015.

[46] F. Baneyx and M. Mujacic, "Recombinant protein folding and misfolding in *Escherichia coli*," *Nat. Biotechnol.*, vol. 22, no. 11, pp. 1399–1408, 2004, doi: 10.1038/nbt1029.

[47] D. G. Kang *et al.*, "Coexpression of molecular chaperone enhances activity and export of organophosphorus hydrolase in *Escherichia coli*," *Biotechnol. Prog.*, vol. 28, no. 4, pp. 925–930, Jul. 2012, doi: 10.1002/BTPR.1556.

[48] A. C. Fisher *et al.*, "Exploration of twin-arginine translocation for expression and purification of correctly folded proteins in *Escherichia coli*," *Microb. Biotechnol.*, vol. 1, no. 5, pp. 403–415, Sep. 2008, doi: 10.1111/J.1751-7915.2008.00041.X.

[49] J. D. Thomas, R. A. Daniel, J. Errington, and C. Robinson, "Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli*," *Mol. Microbiol.*, vol. 39, no. 1, pp. 47–53, 2001, doi: 10.1046/J.1365-2958.2001.02253.X.

[50] M. P. DeLisat, D. Tullman, and G. Georgiou, "Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 10, pp. 6115–6120, May 2003, doi: 10.1073/PNAS.0937838100.

[51] R. Freudl, "Signal peptides for recombinant protein secretion in bacterial expression systems," *Microbial Cell Factories*. 2018, doi: 10.1186/s12934-018-0901-3.

[52] K. M. Frain, J. M. van Dijl, and C. Robinson, "The Twin-Arginine Pathway for Protein Secretion," *EcoSal Plus*, vol. 8, no. 2, Feb. 2019, doi: 10.1128/ECOSALPLUS.ESP-0040-2018/ASSET/76F69F59-936E-4500-AF2C-A419D40C1ED9/ASSETS/GRAPHIC/PSIB-0023-2018_FIG_003.GIF.

[53] M. N. Taw, M. Li, D. Kim, M. A. Rocco, D. Waraho-Zhmayev, and M. P. DeLisa, "Engineering a Supersecreting Strain of*Escherichia coli*by Directed Coevolution of the Multiprotein Tat Translocation Machinery," *ACS Synth. Biol.*, vol. 10, no. 11, pp. 2947–2958, Nov. 2021, doi: 10.1021/ACSSYNBIO.1C00183/SUPPL_FILE/SB1C00183_SI_001.PDF.

[54] J. A. Pogliano and J. Beckwith, "SecD and SecF facilitate protein export in *Escherichia coli*.," *EMBO J.*, 1994, doi: 10.1002/j.1460-2075.1994.tb06293.x.

[55] A. Tsirigotaki, J. De Geyter, N. Šoštarić, A. Economou, and S. Karamanou, "Protein export through the bacterial Sec pathway," *Nature Reviews*

Microbiology, vol. 15, no. 1. Nature Publishing Group, pp. 21–36, Jan. 01, 2017, doi: 10.1038/nrmicro.2016.161.

[56] J. M. Crane and L. L. Randall, "The Sec System: Protein Export in *Escherichia coli*," *EcoSal Plus*, 2017, doi: 10.1128/ecosalplus.esp-0002-2017.

[57] E. Papanikou *et al.*, "Identification of the preprotein binding domain of SecA," *J. Biol. Chem.*, vol. 280, no. 52, pp. 43209–43217, Dec. 2005, doi: 10.1074/JBC.M509990200.

[58] A. J. M. Driessen, "SecB, a molecular chaperone with two faces," *Trends Microbiol.*, vol. 9, no. 5, pp. 193–196, May 2001, doi: 10.1016/S0966-842X(01)01980-1.

[59] L. McFarland, O. Francetic, and C. A. Kumamoto, "A mutation of *Escherichia coli* SecA protein that partially compensates for the absence of SecB," *J. Bacteriol.*, 1993, doi: 10.1128/jb.175.8.2255-2262.1993.

[60] J. Kim, J. Luirink, and D. A. Kendall, "SecB dependence of an exported protein is a continuum influenced by the characteristics of the signal peptide or early mature region," *J. Bacteriol.*, vol. 182, no. 14, pp. 4108–4112, Jul. 2000, doi: 10.1128/JB.182.14.4108-4112.2000.

[61] B. W. Bauer, T. Shemesh, Y. Chen, and T. A. Rapoport, "A 'push and slide' mechanism allows sequence-insensitive translocation of secretory proteins by the SecA ATPase," *Cell*, vol. 157, no. 6, pp. 1416–1429, Jun. 2014, doi: 10.1016/J.CELL.2014.03.063.

[62] W. J. Allen *et al.*, "Two-way communication between SecY and SecA suggests a brownian ratchet mechanism for protein translocation," *Elife*, vol. 5, no. MAY2016, May 2016, doi: 10.7554/ELIFE.15598.

[63] C. Jiang, M. Wynne, and D. Huber, "How Quality Control Systems AID Sec-Dependent Protein Translocation," *Front. Mol. Biosci.*, vol. 8, Apr. 2021, doi: 10.3389/FMOLB.2021.669376.

[64] J. Wild, E. Altman, T. Yura, and C. A. Gross, "DnaK and DnaJ heat shock proteins participate in protein export in *Escherichia coli*," *Genes Dev.*, vol. 6, no. 7, pp. 1165–1172, 1992, doi: 10.1101/GAD.6.7.1165.

[65] A. R. Duguay and T. J. Silhavy, "Quality control in the bacterial periplasm," *Biochim. Biophys. Acta-Molecular Cell Res.*, vol. 1694, no. 1–3, pp. 121–134, 2004, doi: 10.1016/j.bbamcr.2004.04.012.

[66] M. Merdanovic, T. Clausen, M. Kaiser, R. Huber, and M. Ehrmann, "Protein Quality Control in the Bacterial Periplasm," *Annu. Rev. Microbiol. Vol 65*, vol. 65, pp. 149-+, 2011, doi: 10.1146/annurev-micro-090110-102925.

[67] J. De Geyter, A. Tsirigotaki, G. Orfanoudaki, V. Zorzini, A. Economou, and S. Karamanou, "Protein folding in the cell envelope of *Escherichia coli*," *Nat. Microbiol.*, 2016, doi: 10.1038/nmicrobiol.2016.107.

[68] F. Stull, J.-M. Betton, and J. C. A. Bardwell, "Periplasmic Chaperones and Prolyl Isomerases," *EcoSal Plus*, vol. 8, no. 1, Feb. 2018, doi: 10.1128/ECOSALPLUS.ESP-0005-2018.

[69] H. Kim, K. Wu, and C. Lee, "Stress-Responsive Periplasmic Chaperones in Bacteria," *Front. Mol. Biosci.*, vol. 8, p. 678697, May 2021, doi: 10.3389/FMOLB.2021.678697.

[70] C. W. Lennon *et al.*, "Folding optimization in vivo uncovers new

chaperones," *J. Mol. Biol.*, vol. 427, no. 18, p. 2983, Sep. 2015, doi: 10.1016/J.JMB.2015.05.013.

[71] G. Richarme and T. D. Caldas, "Chaperone Properties of the Bacterial Periplasmic Substrate-binding Proteins," *J. Biol. Chem.*, vol. 272, no. 25, pp. 15607–15612, Jun. 1997, doi: 10.1074/JBC.272.25.15607.

[72] B. Manta, D. Boyd, and M. Berkmen, " Disulfide Bond Formation in the Periplasm of *Escherichia coli* ," *EcoSal Plus*, vol. 8, no. 2, Feb. 2019, doi: 10.1128/ECOSALPLUS.ESP-0012-2018.

[73] K. Inaba, "Disulfide bond formation system in *Escherichia coli*," *Journal of Biochemistry*, vol. 146, no. 5. pp. 591–597, Nov. 2009, doi: 10.1093/jb/mvp102.

[74] J. C. A. Bardwell, K. McGovern, and J. Beckwith, "Identification of a protein required for disulfide bond formation in vivo," *Cell*, vol. 67, no. 3, pp. 581–589, Nov. 1991, doi: 10.1016/0092-8674(91)90532-4.

[75] J. C. A. Bardwell, J. O. Lee, G. Jander, N. Martin, D. Belin, and J. Beckwith, "A pathway for disulfide bond formation in vivo," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 3, pp. 1038–1042, Feb. 1993, doi: 10.1073/PNAS.90.3.1038.

[76] K. M. Bocian-Ostrzycka, M. J. Grzeszczuk, L. Dziewit, and E. K. Jagusztyn-Krynicka, "Diversity of the epsilonproteobacteria Dsb (disulfide bond) systems," *Front. Microbiol.*, vol. 6, no. JUN, p. 570, 2015, doi: 10.3389/FMICB.2015.00570/ABSTRACT.

[77] P. E. Rouvière, A. De Las Peñas, J. Mecsas, C. Z. Lu, K. E. Rudd, and C. A. Gross, "rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*.," *EMBO J.*, 1995, doi: 10.1002/j.1460-2075.1995.tb07084.x.

[78] S. Bury-Mone *et al.*, "Global Analysis of Extracytoplasmic Stress Signaling in *Escherichia coli*," *Plos Genet.*, vol. 5, no. 9, 2009, doi: 10.1371/journal.pgen.1000651.

[79] J. Mecsas, P. E. Rouviere, J. W. Erickson, T. J. Donohue, and C. A. Gross, "The activity of σ(E), an *Escherichia coli* heat-inducible σ-factor, is modulated by expression of outer membrane proteins," *Genes Dev.*, 1993, doi: 10.1101/gad.7.12b.2618.

[80] N. P. Walsh, B. M. Alba, B. Bose, C. A. Gross, and R. T. Sauer, "OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain," *Cell*, 2003, doi: 10.1016/S0092-8674(03)00203-4.

[81] G. Rowley, M. Spector, J. Kormanec, and M. Roberts, "Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens," *Nat. Rev. Microbiol.*, vol. 4, no. 5, pp. 383–394, 2006, doi: 10.1038/nrmicro1394.

[82] A. M. Mitchell and T. J. Silhavy, "Envelope stress responses: balancing damage repair and toxicity," *Nature Reviews Microbiology*. 2019, doi: 10.1038/s41579-019-0199-0.

[83] S. Raina, D. Missiakas, and C. Georgopoulos, "The rpoE gene encoding the sigma E (sigma 24) heat shock sigma factor of *Escherichia coli*.," *EMBO J.*, 1995, doi: 10.1002/j.1460-2075.1995.tb07085.x.

[84] T. L. Raivio, S. K. D. Leblanc, and N. L. Price, "The *Escherichia coli* Cpx Envelope Stress Response Regulates Genes of Diverse Function That

Impact Antibiotic Resistance and Membrane Integrity," *J. Bacteriol.*, vol. 195, no. 12, p. 2755, 2013, doi: 10.1128/JB.00105-13.

[85] N. Joly *et al.*, "Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology," *Fems Microbiol. Rev.*, vol. 34, no. 5, pp. 797–827, 2010, doi: 10.1111/j.1574-6976.2010.00240.x.

[86] G. Jovanovic, C. Engl, A. J. Mayhew, P. C. Burrows, and M. Buck, "Properties of the phage-shock-protein (Psp) regulatory complex that govern signal transduction and induction of the Psp response in *Escherichia coli*," *Microbiology-Sgm*, vol. 156, pp. 2920–2932, 2010, doi: 10.1099/mic.0.040055-0.

[87] R. Kobayashi, T. Suzuki, and M. Yoshida, "*Escherichia coli* phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes," *Mol. Microbiol.*, vol. 66, no. 1, pp. 100–109, Oct. 2007, doi: 10.1111/J.1365-2958.2007.05893.X.

[88] M. Kleerebezem, W. Crielaard, and J. Tommassen, "Involvement of stress protein PspA (phage shock protein A) of *Escherichia coli* in maintenance of the protonmotive force under stress conditions.," *EMBO J.*, 1996, doi: 10.1002/j.1460-2075.1996.tb00344.x.

[89] S. Schlegel, E. Rujas, A. J. Ytterberg, R. A. Zubarev, J. Luirink, and J.-W. de Gier, "Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels," *Microb. Cell Fact.*, vol. 12, no. 1, 2013, doi: 10.1186/1475-2859-12-24.

[90] T. Baumgarten, A. J. Ytterberg, R. A. Zubarev, and J. W. de Gier, "Optimizing Recombinant Protein Production in the *Escherichia coli* Periplasm Alleviates Stress," *Appl. Environ. Microbiol.*, vol. 84, no. 12, 2018, doi: 10.1128/aem.00270-18.

[91] J. M. Newton, D. Schofield, J. Vlahopoulou, and Y. Zhou, "Detecting cell lysis using viscosity monitoring in *E. coli* fermentation to prevent product loss," *Biotechnol. Prog.*, 2016, doi: 10.1002/btpr.2292.

[92] J. H. Choi and S. Y. Lee, "Secretory and extracellular production of recombinant proteins using *Escherichia coli*," *Appl. Microbiol. Biotechnol.*, vol. 64, no. 5, pp. 625–635, 2004, doi: 10.1007/s00253-004-1559-9.

[93] W. Chan *et al.*, "Mutational effects on inclusion body formation in the periplasmic expression of the immunoglobulin VL domain REI," *Fold. Des.*, 1996, doi: 10.1016/S1359-0278(96)00017-X.

[94] J. P. Arie, M. Miot, N. Sassoon, and J. M. Betton, "Formation of active inclusion bodies in the periplasm of *Escherichia coli*," *Mol. Microbiol.*, vol. 62, no. 2, pp. 427–437, 2006, doi: 10.1111/j.1365-2958.2006.05394.x.

[95] N. Sriubolmas, W. Panbangred, S. Sriurairatana, and V. Meevootisom, "Localization and characterization of inclusion bodies in recombinant *Escherichia coli* cells overproducing penicillin G acylase," *Appl. Microbiol. Biotechnol.*, vol. 47, no. 4, pp. 373–378, 1997, doi: 10.1007/S002530050943.

[96] S. Hunke and J. M. Betton, "Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*," *Mol. Microbiol.*, vol. 50, no. 5, pp. 1579–1589, Dec. 2003, doi: 10.1046/J.1365-2958.2003.03785.X.

[97] S. Wagner *et al.*, "Consequences of membrane protein overexpression in *Escherichia coli*," *Mol. Cell. Proteomics*, vol. 6, no. 9, pp. 1527–1550, 2007, doi: 10.1074/mcp.M600431-MCP200.

[98] A. Karyolaimos *et al.*, "Enhancing Recombinant Protein Yields in the *E. coli* Periplasm by Combining Signal Peptide and Production Rate Screening," *Front. Microbiol.*, vol. 10, p. 1511, Jul. 2019, doi: 10.3389/fmicb.2019.01511.

[99] A. Gawin, H. Ertesvåg, S. A. H. Hansen, J. Malmo, and T. Brautaset, "Translational regulation of periplasmic folding assistants and proteases as a valuable strategy to improve production of translocated recombinant proteins in *Escherichia coli*," *BMC Biotechnol.*, vol. 20, no. 1, p. 1DUMM, May 2020, doi: 10.1186/S12896-020-00615-0.

[100] H. Sonoda, Y. Kumada, T. Katsuda, and H. Yamaji, "Effects of cytoplasmic and periplasmic chaperones on secretory production of single-chain Fv antibody in *Escherichia coli*," *J. Biosci. Bioeng.*, vol. 111, no. 4, pp. 465–470, Apr. 2011, doi: 10.1016/J.JBIOSC.2010.12.015.

[101] T. Makino, G. Skretas, T. H. Kang, and G. Georgiou, "Comprehensive Engineering of *Escherichia coli* for Enhanced Expression of IgG Antibodies," *Metab. Eng.*, vol. 13, no. 2, p. 241, Mar. 2011, doi: 10.1016/J.YMBEN.2010.11.002.

[102] A. Karyolaimos and J. W. de Gier, "Strategies to Enhance Periplasmic Recombinant Protein Production Yields in *Escherichia coli*," *Front. Bioeng. Biotechnol.*, vol. 9, Dec. 2021, doi: 10.3389/FBIOE.2021.797334.

[103] A. de Marco, "Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*," *Microbial Cell Factories*, vol. 8. p. 26, May 14, 2009, doi: 10.1186/1475-2859-8-26.

[104] C. F. Schierle, M. Berkmen, D. Huber, C. Kumamoto, D. Boyd, and J. Beckwith, "The DsbA Signal Sequence Directs Efficient, Cotranslational Export of Passenger Proteins to the *Escherichia coli* Periplasm via the Signal Recognition Particle Pathway," *J. Bacteriol.*, vol. 185, no. 19, p. 5706, Oct. 2003, doi: 10.1128/JB.185.19.5706-5713.2003.

[105] K. O. Low, N. M. Mahadi, and R. M. Illias, "Optimisation of signal peptide for recombinant protein secretion in bacterial hosts," *Appl. Microbiol. Biotechnol.*, vol. 97, no. 9, pp. 3811–3826, May 2013, doi: 10.1007/S00253-013-4831-Z/TABLES/3.

[106] A. Karyolaimos *et al.*, "*Escherichia coli* Can Adapt Its Protein Translocation Machinery for Enhanced Periplasmic Recombinant Protein Production," *Front. Bioeng. Biotechnol.*, vol. 7, Jan. 2020, doi: 10.3389/fbioe.2019.00465.

[107] A. J. Ytterberg, R. A. Zubarev, and T. Baumgarten, "Posttranslational targeting of a recombinant protein promotes its efficient secretion into the *Escherichia coli* periplasm," *Appl. Environ. Microbiol.*, 2019, doi: 10.1128/AEM.00671-19.

[108] Y. J. Lee and K. J. Jeong, "Enhanced production of antibody fragment via SRP pathway engineering in *Escherichia coli*," *Biotechnol. Bioprocess Eng.*, vol. 18, no. 4, pp. 751–758, Jul. 2013, doi: 10.1007/s12257-013-0111-0.

[109] J. Goldstein, S. Lehnhardt, and M. Inouye, "Enhancement of protein translocation across the membrane by specific mutations in the

hydrophobic region of the signal peptide.," *J. Bacteriol.*, vol. 172, no. 3, p. 1225, 1990, doi: 10.1128/JB.172.3.1225-1231.1990.

[110] B. K. Klein, J. O. Polazzi, C. S. Devine, S. H. Rangwala, and P. O. Olins, "Effects of signal peptide changes on the secretion of bovine somatotropin (bST) from *Escherichia coli*," *Protein Eng.*, vol. 5, no. 6, pp. 511–517, Sep. 1992, doi: 10.1093/PROTEIN/5.6.511.

[111] K. Morioka-Fujimoto, R. Marumoto, and T. Fukuda, "Modified enterotoxin signal sequences increase secretion level of the recombinant human epidermal growth factor in *Escherichia coli*," *J. Biol. Chem.*, vol. 266, no. 3, pp. 1728–1732, Jan. 1991, doi: 10.1016/S0021-9258(18)52356-5.

[112] A. Hjelm *et al.*, "Tailoring *Escherichia coli* for the <scp>l</scp> -Rhamnose P BAD Promoter-Based Production of Membrane and Secretory Proteins," *ACS Synth. Biol.*, vol. 6, no. 6, Jun. 2017, doi: 10.1021/acssynbio.6b00321.

[113] H. Sletta *et al.*, "Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in *Escherichia coli*," *Appl. Environ. Microbiol.*, vol. 70, no. 12, pp. 7033–7039, Dec. 2004, doi: 10.1128/AEM.70.12.7033-7039.2004.

[114] K. J. Jeong and S. Y. Lee, "High-level production of human leptin by fed-batch cultivation of recombinant *Escherichia coli* and its purification," *Appl. Environ. Microbiol.*, 1999.

[115] F. J. M. Mergulhão *et al.*, "Medium and copy number effects on the secretion of human proinsulin in *Escherichia coli* using the universal stress promoters uspA and uspB," *Appl. Microbiol. Biotechnol.*, vol. 61, no. 5–6, pp. 495–501, 2003, doi: 10.1007/S00253-003-1232-8.

[116] L. G. Horga *et al.*, "Tuning recombinant protein expression to match secretion capacity," *Microb. Cell Fact.*, vol. 17, no. 1, pp. 1–18, Dec. 2018, doi: 10.1186/s12934-018-1047-z.

[117] L. C. Simmons and D. G. Yansura, "Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*," *Nat. Biotechnol. 1996 145*, vol. 14, no. 5, pp. 629–634, 1996, doi: 10.1038/nbt0596-629.

[118] H. Waegeman and W. Soetaert, "Increasing recombinant protein production in *Escherichia coli* through metabolic and genetic engineering," *Journal of Industrial Microbiology and Biotechnology*. 2011, doi: 10.1007/s10295-011-1034-4.

[119] T. Makino, G. Skretas, and G. Georgiou, "Strain engineering for improved expression of recombinant proteins in bacteria," *Microb. Cell Fact.*, vol. 10, 2011, doi: 10.1186/1475-2859-10-32.

[120] S. Mahalik, A. K. Sharma, and K. J. Mukherjee, "Genome engineering for improved recombinant protein expression in *Escherichia coli*," *Microb. Cell Fact.*, vol. 13, 2014, doi: 10.1186/s12934-014-0177-1.

[121] H. P. Sørensen and K. K. Mortensen, "Advanced genetic strategies for recombinant protein expression in *Escherichia coli*," *J. Biotechnol.*, vol. 115, no. 2, pp. 113–128, Jan. 2005, doi: 10.1016/J.JBIOTEC.2004.08.004.

[122] G. Georgiou and L. Segatori, "Preparative expression of secreted proteins in bacteria: status report and future prospects," *Curr. Opin. Biotechnol.*, vol. 16, no. 5, pp. 538–545, 2005, doi: 10.1016/j.copbio.2005.07.008.

[123] A. S. L. Hansen, R. M. Lennen, N. Sonnenschein, and M. J. Herrgård,

"Systems biology solutions for biochemical production challenges," *Curr. Opin. Biotechnol.*, vol. 45, pp. 85–91, Jun. 2017, doi: 10.1016/J.COPBIO.2016.11.018.

[124] T. U. Chae, S. Y. Choi, J. W. Kim, Y. S. Ko, and S. Y. Lee, "Recent advances in systems metabolic engineering tools and strategies," *Curr. Opin. Biotechnol.*, vol. 47, pp. 67–82, Oct. 2017, doi: 10.1016/J.COPBIO.2017.06.007.

[125] Z. Ignatova, A. Mahsunah, M. Georgieva, and V. Kasche, "Improvement of posttranslational bottlenecks in the production of penicillin amidase in recombinant *Escherichia coli* strains," *Appl. Environ. Microbiol.*, 2003, doi: 10.1128/AEM.69.2.1237-1245.2003.

[126] S. Farajnia, V. Ghorbanzadeh, and H. Dariushnejad, "Effect of Molecular Chaperone on the Soluble Expression of Recombinant Fab Fragment in *E. coli*," *Int. J. Pept. Res. Ther.*, vol. 26, no. 1, pp. 251–258, Mar. 2020, doi: 10.1007/S10989-019-09833-3/FIGURES/7.

[127] X. Hu, L. O'Hara, S. White, E. Magner, M. Kane, and J. Gerard Wall, "Optimisation of production of a domoic acid-binding scFv antibody fragment in *Escherichia coli* using molecular chaperones and functional immobilisation on a mesoporous silicate support," *Protein Expr. Purif.*, vol. 52, no. 1, pp. 194–201, Mar. 2007, doi: 10.1016/J.PEP.2006.08.009.

[128] H. Dariushnejad, S. Farajnia, N. Zarghami, M. Aria, and A. Tanomand, "Effect of DnaK/DnaJ/GrpE and DsbC Chaperons on Periplasmic Expression of Fab Antibody by *E. coli* SEC Pathway," *Int. J. Pept. Res. Ther.*, vol. 25, no. 1, pp. 67–74, Mar. 2019, doi: 10.1007/S10989-017-9637-X.

[129] J. Pérez-Pérez, C. Martínez-Caja, J. L. Barbero, and J. Gutiérrez, "DnaK/DnaJ supplementation improves the periplasmic production of human granulocyte-colony stimulating factor in *Escherichia coli*," *Biochem. Biophys. Res. Commun.*, vol. 210, no. 2, pp. 524–529, 1995, doi: 10.1006/BBRC.1995.1691.

[130] J. De Geyter, A. G. Portaliou, B. Srinivasu, S. Krishnamurthy, A. Economou, and S. Karamanou, "Trigger factor is a bona fide secretory pathway chaperone that interacts with SecB and the translocase," *EMBO Rep.*, vol. 21, no. 6, Jun. 2020, doi: 10.15252/EMBR.201949054.

[131] J. M. Puertas, B. L. Nannenga, K. T. Dornfeld, J. M. Betton, and F. Baneyx, "Enhancing the secretory yields of leech carboxypeptidase inhibitor in *Escherichia coli*: Influence of trigger factor and signal recognition particle," *Protein Expr. Purif.*, vol. 74, no. 1, pp. 122–128, Nov. 2010, doi: 10.1016/J.PEP.2010.06.008.

[132] C. F. R. O. Matos *et al.*, "High-yield export of a native heterologous protein to the periplasm byMatos, C. F. R. O., Branston, S. D., Albiniak, A., Dhanoya, A., Freedman, R. B., Keshavarz-Moore, E., & Robinson, C. (2012). High-yield export of a native heterologous protein to the pe," *Biotechnol. Bioeng.*, 2012, doi: 10.1002/bit.24535.

[133] D. F. Browning, K. L. Richards, A. R. Peswani, J. Roobol, S. J. W. Busby, and C. Robinson, "*Escherichia coli* 'TatExpress' strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway," *Biotechnol. Bioeng.*, 2017, doi: 10.1002/bit.26434.

[134] J. Pérez-Pérez, G. Márquez, J. L. Barbero, and J. Gutiérrez, "Increasing the

Efficiency of Protein Export in *Escherichia coli*," *Bio/Technology 1994 122*, vol. 12, no. 2, pp. 178–180, 1994, doi: 10.1038/nbt0294-178.

[135] K. J. Jeong and S. Y. Lee, "Secretory production of human leptin in *Escherichia coli*," *Biotechnol. Bioeng.*, 2000, doi: 10.1002/(SICI)1097-0290(20000220)67:4<398::AID-BIT3>3.0.CO;2-Y.

[136] J. C. Joly, W. S. Leung, and J. R. Swartz, "Overexpression of *Escherichia coli* oxidoreductases increases recombinant insulin-like growth factor-I accumulation," *Proc. Natl. Acad. Sci.*, 2002, doi: 10.1073/pnas.95.6.2773.

[137] M. Ellis, P. Patel, M. Edon, W. Ramage, R. Dickinson, and D. P. Humphreys, "Development of a high yielding *E. coli* periplasmic expression system for the production of humanized Fab' fragments," *Biotechnol. Prog.*, 2017, doi: 10.1002/btpr.2393.

[138] H. Bothmann and A. Plückthun, "Selection for a periplasmic factor improving phage display and functional periplasmic expression," *Nat. Biotechnol.*, 1998, doi: 10.1038/nbt0498-376.

[139] N. Narayanan and C. P. Chou, "Physiological improvement to enhance *Escherichia coli* cell-surface display via reducing extracytoplasmic stress," *Biotechnol. Prog.*, 2008, doi: 10.1021/bp0702121.

[140] C. Mavrangelos *et al.*, "Increased yield and activity of soluble single-chain antibody fragments by combining high-level expression and the Skp periplasmic chaperonin," *Protein Expr. Purif.*, vol. 23, no. 2, pp. 289–295, 2001, doi: 10.1006/PREP.2001.1506.

[141] H. Bothmann and A. Plückthun, "The periplasmic *Escherichia coli* peptidylprolyl cis,trans-isomerase FkpA. I. Increased functional expression of antibody fragments with and without cis-prolines," *J. Biol. Chem.*, vol. 275, no. 22, pp. 17100–17105, Jun. 2000, doi: 10.1074/JBC.M910233199.

[142] C. P. Chou, "Engineering cell physiology to enhance recombinant protein production in *Escherichia coli*," *Applied Microbiology and Biotechnology*. 2007, doi: 10.1007/s00253-007-1039-0.

[143] R. Levy *et al.*, "Enhancement of antibody fragment secretion into the *Escherichia coli* periplasm by co-expression with the peptidyl prolyl isomerase, FkpA, in the cytoplasm," *J. Immunol. Methods*, 2013, doi: 10.1016/j.jim.2013.04.010.

[144] C. Chen *et al.*, "High-level accumulation of a recombinant antibody fragment in the periplasm of *Escherichia coli* requires a triple-mutant (degP prc spr) host strain," *Biotechnol. Bioeng.*, vol. 85, no. 5, pp. 463–474, 2004, doi: 10.1002/bit.20014.

[145] H. J. Meerman and G. Georgiou, "Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins," *Biotechnology. (N. Y).*, vol. 12, no. 11, pp. 1107–1110, 1994, doi: 10.1038/NBT1194-1107.

[146] I. S. Aldor *et al.*, "Proteomic profiling of recombinant *Escherichia coli* in high-cell-density fermentations for improved production of an antibody fragment biopharmaceutical," *Appl. Environ. Microbiol.*, vol. 71, no. 4, pp. 1717–1728, 2005, doi: 10.1128/aem.71.4.1717-1728.2005.

[147] Y. Y. Wang *et al.*, "Enhancement of Excretory Production of an Exoglucanase from *Escherichia coli* with Phage Shock Protein A (PspA)

Overexpression," *J. Microbiol. Biotechnol.*, vol. 21, no. 6, pp. 637–645, 2011, doi: 10.4014/jmb.1101.01036.

[148] K. M. Champion, J. C. Nishihara, J. C. Joly, and D. Arnott, "Similarity of the *Escherichia coli* proteome upon completion of different biopharmaceutical fermentation processes," *Proteomics*, vol. 1, no. 9, pp. 1133–1148, 2001, doi: 10.1002/1615-9861(200109)1:9<1133::aid-prot1133>3.3.co;2-j.

[149] M. P. DeLisa, P. Lee, T. Palmer, and G. Georgiou, "Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway," *J. Bacteriol.*, vol. 186, no. 2, pp. 366–373, 2004, doi: 10.1128/jb.186.2.366-373.2004.

[150] W. E. Bentley, N. Mirjalili, D. C. Andersen, R. H. Davis, and D. S. Kompala, "Plasmid encoded protein: The principal factor in the metabolic burden associated with recombinant bacteria," *Biotechnol. Bioeng.*, 1990, doi: 10.1002/bit.260350704.

[151] H. A. George *et al.*, "Physiological effects of TGFα-PE40 expression in recombinant *Escherichia coli* JM109," *Biotechnol. Bioeng.*, 1992, doi: 10.1002/bit.260400314.

[152] L. Andersson, S. Yang, P. Neubauer, and S. O. Enfors, "Impact of plasmid presence and induction on cellular responses in fed batch cultures of *Escherichia coli*," *J. Biotechnol.*, 1996, doi: 10.1016/0168-1656(96)00004-1.

[153] H. Dong, L. Nilsson, and C. G. Kurland, "Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction," *J. Bacteriol.*, 1995, doi: 10.1128/jb.177.6.1497-1504.1995.

[154] C. G. Kurland and H. Dong, "Bacterial growth inhibition by overproduction of protein," *Molecular Microbiology*. 1996, doi: 10.1046/j.1365-2958.1996.5901313.x.

[155] J. C. Diaz Ricci and M. E. Hernández, "Plasmid effects on *Escherichia coli* metabolism," *Crit. Rev. Biotechnol.*, vol. 20, no. 2, pp. 79–108, 2000, doi: 10.1080/07388550008984167.

[156] B. R. Glick, "Metabolic load and heterologous gene expression," *Biotechnology Advances*. 1995, doi: 10.1016/0734-9750(95)00004-A.

[157] F. Hoffmann and U. Rinas, "Stress induced by recombinant protein production in *Escherichia coli*.," *Adv. Biochem. Eng. Biotechnol.*, 2004.

[158] S. A. Goff and A. L. Goldberg, "Production of abnormal proteins in *E. coli* stimulates transcription of ion and other heat shock genes," *Cell*, 1985, doi: 10.1016/S0092-8674(85)80031-3.

[159] S. W. Harcum and W. E. Bentley, "Response dynamics of 26-, 34-, 39-, 54-, and 80-kDa proteases in induced cultures of recombinant *Escherichia coli*," *Biotechnol. Bioeng.*, 1993, doi: 10.1002/bit.260420602.

[160] J. Becker and C. Wittmann, "From systems biology to metabolically engineered cells — an omics perspective on the development of industrial microbes," *Curr. Opin. Microbiol.*, vol. 45, pp. 180–188, Oct. 2018, doi: 10.1016/J.MIB.2018.06.001.

[161] A. Landels, C. Evans, J. Noirel, and P. C. Wright, "Advances in proteomics for production strain analysis," *Curr. Opin. Biotechnol.*, vol. 35, pp. 111–117, 2015, doi: 10.1016/j.copbio.2015.05.001.

[162] J. M. Monk *et al.*, "iML1515, a knowledgebase that computes *Escherichia coli* traits," *Nat. Biotechnol.*, 2017, doi: 10.1038/nbt.3956.

[163] P. D. Karp, "The EcoCyc and MetaCyc databases," *Nucleic Acids Res.*, 2000, doi: 10.1093/nar/28.1.56.

[164] E. Rowe, B. O. Palsson, and Z. A. King, "Escher-FBA: a web application for interactive flux balance analysis," *BMC Syst. Biol.*, 2018, doi: 10.1186/s12918-018-0607-5.

[165] W. W. Zhang, F. Li, and L. Nie, "Integrating multiple 'omics' analysis for microbial biology: application and methodologies," *Microbiology-Sgm*, vol. 156, pp. 287–301, 2010, doi: 10.1099/mic.0.034793-0.

[166] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray," *Science (80-.).*, 1995, doi: 10.1126/science.270.5235.467.

[167] T. Casneuf, Y. Van de Peer, and W. Huber, "In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation," *BMC Bioinformatics*, 2007, doi: 10.1186/1471-2105-8-461.

[168] R. T. Gill, J. J. Valdes, and W. E. Bentley, "A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in Escherichiaa coli," *Metab. Eng.*, 2000, doi: 10.1006/mben.2000.0148.

[169] M. K. Oh and J. C. Liao, "DNA microarray detection of metabolic responses to protein overproduction in *Escherichia coli*," *Metab. Eng.*, 2000, doi: 10.1006/mben.2000.0149.

[170] R. T. Gill, M. P. DeLisa, J. J. Valdes, and W. E. Bentley, "Genomic analysis of high-cell-density recombinant *Escherichia coli* fermentation and 'cell conditioning' for improved recombinant protein yield," *Biotechnol. Bioeng.*, 2001, doi: 10.1002/1097-0290(20010105)72:1<85::AID-BIT12>3.0.CO;2-H.

[171] P. Muir *et al.*, "The real cost of sequencing: scaling computation to keep pace with data generation," *Genome Biol.*, vol. 17, no. 1, p. 53, Mar. 2016, doi: 10.1186/S13059-016-0917-0.

[172] B. J. Haas, M. Chin, C. Nusbaum, B. W. Birren, and J. Livny, "How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?," *BMC Genomics*, vol. 13, no. 1, pp. 1–11, Dec. 2012, doi: 10.1186/1471-2164-13-734/FIGURES/6.

[173] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics.," *Nat. Rev. Genet.*, 2009, doi: 10.1038/nrg2484.

[174] B. E. Slatko, A. F. Gardner, and F. M. Ausubel, "Overview of Next Generation Sequencing Technologies," *Curr. Protoc. Mol. Biol.*, vol. 122, no. 1, p. e59, Apr. 2018, doi: 10.1002/CPMB.59.

[175] C. Delahaye and J. Nicolas, "Sequencing DNA with nanopores: Troubles and biases," *PLoS One*, vol. 16, no. 10, p. e0257521, Oct. 2021, doi: 10.1371/JOURNAL.PONE.0257521.

[176] Y. Wang, Y. Zhao, A. Bollas, Y. Wang, and K. F. Au, "Nanopore sequencing technology, bioinformatics and applications," *Nat. Biotechnol. 2021 3911*, vol. 39, no. 11, pp. 1348–1365, Nov. 2021, doi: 10.1038/s41587-021-01108-x.

[177] T. Maier, M. Güell, and L. Serrano, "Correlation of mRNA and protein in complex biological samples.," *FEBS Lett.*, 2009, doi: 10.1016/j.febslet.2009.10.036.

[178] R. A. Zubarev, "The challenge of the proteome dynamic range and its implications for in-depth proteomics," *Proteomics*, vol. 13, no. 5, pp. 723–726, Mar. 2013, doi: 10.1002/PMIC.201200451.

[179] S. Magdeldin *et al.*, "Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis," *Clin. Proteomics*, vol. 11, no. 1, pp. 1–10, Apr. 2014, doi: 10.1186/1559-0275-11-16/FIGURES/5.

[180] R. P. Tonge *et al.*, "Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology," *Proteomics*, 2001, doi: 10.1002/1615-9861(200103)1:3<377::AID-PROT377>3.0.CO;2-6.

[181] Y. Yamaguchi, T. Tomoyasu, A. Takaya, M. Morioka, and T. Yamamoto, "Effects of disruption of heat shock genes on susceptibility of *Escherichia coli* to fluoroquinolones," *BMC Microbiol.*, vol. 3, no. 1, pp. 1–8, Aug. 2003, doi: 10.1186/1471-2180-3-16/TABLES/1.

[182] C. Evans *et al.*, "An insight into iTRAQ: where do we stand now?," *Anal. Bioanal. Chem.*, vol. 404, no. 4, pp. 1011–1027, 2012, doi: 10.1007/s00216-012-5918-6.

[183] D. A. Megger, T. Bracht, H. E. Meyer, and B. Sitek, "Label-free quantification in clinical proteomics," *Biochim. Biophys. Acta - Proteins Proteomics*, vol. 1834, no. 8, pp. 1581–1590, Aug. 2013, doi: 10.1016/J.BBAPAP.2013.04.001.

[184] H. Wang, S. Alvarez, and L. M. Hicks, "Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two chlamydomonas reinhardtii strains of interest for biofuels engineering," *J. Proteome Res.*, 2012, doi: 10.1021/pr2008225.

[185] Z. Li, R. M. Adams, K. Chourey, G. B. Hurst, R. L. Hettich, and C. Pan, "Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ orbitrap velos," 2012, doi: 10.1021/pr200748h.

[186] J. Li *et al.*, "TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing," *J. Proteome Res.*, vol. 20, no. 5, pp. 2964–2972, May 2021, doi: 10.1021/ACS.JPROTEOME.1C00168.

[187] P. Mertins *et al.*, "iTRAQ Labeling is Superior to mTRAQ for Quantitative Global Proteomics and Phosphoproteomics," *Mol. Cell. Proteomics*, vol. 11, no. 6, 2012, doi: 10.1074/mcp.M111.014423.

[188] A. S. Sandberg, R. M. M. Branca, J. Lehtiö, and J. Forshed, "Quantitative accuracy in mass spectrometry based proteomics of complex samples: The impact of labeling and precursor interference," *J. Proteomics*, 2014, doi: 10.1016/j.jprot.2013.10.035.

[189] S. Y. Ow, M. Salim, J. Noirel, C. Evans, I. Rehman, and P. C. Wright, "iTRAQ Underestimation in Simple and Complex Mixtures: 'The Good, the Bad and the Ugly,'" *J. Proteome Res.*, vol. 8, no. 11, pp. 5347–5355, 2009, doi: 10.1021/pr900634c.

[190] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster, "Quantitative

mass spectrometry in proteomics: A critical review," *Anal. Bioanal. Chem.*, 2007, doi: 10.1007/s00216-007-1486-6.

[191] C. D. Wenger *et al.*, "Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging," *Nat. Methods 2011 811*, vol. 8, no. 11, pp. 933–935, Oct. 2011, doi: 10.1038/nmeth.1716.

[192] L. Ting, R. Rad, S. P. Gygi, and W. Haas, "MS3 eliminates ratio distortion in isobaric labeling-based multiplexed quantitative proteomics," *Nat. Methods*, vol. 8, no. 11, p. 937, Nov. 2011, doi: 10.1038/NMETH.1714.

[193] S. Aggarwal and A. K. Yadav, "Dissecting the iTRAQ data analysis," in *Methods in Molecular Biology*, 2016.

[194] T. Sajed *et al.*, "ECMDB 2.0: A richer resource for understanding the biochemistry of *E. coli*," *Nucleic Acids Res.*, vol. 44, no. Database issue, p. D495, 2016, doi: 10.1093/NAR/GKV1060.

[195] W. Lu, X. Su, M. S. Klein, I. A. Lewis, O. Fiehn, and J. D. Rabinowitz, "Metabolite Measurement: Pitfalls to Avoid and Practices to Follow," *Annu. Rev. Biochem.*, vol. 86, p. 277, Jun. 2017, doi: 10.1146/ANNUREV-BIOCHEM-061516-044952.

[196] A. H. M. Emwas, "The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research," *Methods Mol. Biol.*, vol. 1277, pp. 161–193, 2015, doi: 10.1007/978-1-4939-2377-9_13.

[197] K. Segers, S. Declerck, D. Mangelings, Y. Vander Heyden, and A. Van Eeckhaut, "Analytical techniques for metabolomic studies: a review," *https://doi.org/10.4155/bio-2019-0014*, vol. 11, no. 24, pp. 2297–2318, Dec. 2019, doi: 10.4155/BIO-2019-0014.

[198] C. H. Johnson and F. J. Gonzalez, "Challenges and Opportunities of Metabolomics," *J. Cell. Physiol.*, vol. 227, no. 8, p. 2975, Aug. 2012, doi: 10.1002/JCP.24002.

[199] C. Jang, L. Chen, and J. D. Rabinowitz, "Metabolomics and Isotope Tracing," *Cell*, vol. 173, no. 4, pp. 822–837, May 2018, doi: 10.1016/J.CELL.2018.03.055.

[200] S. A. Lesley, J. Graziano, C. Y. Cho, M. W. Knuth, and H. E. Klock, "Gene expression response to misfoldead protein as a screen for soluble recombinant protein," *Protein Eng. Des. Sel.*, 2002, doi: 10.1093/protein/15.2.153.

[201] J. H. Choi, S. J. Lee, S. J. Lee, and S. Y. Lee, "Enhanced production of insulin-like growth factor I fusion protein in *Escherichia coli* by coexpression of the down-regulated genes identified by transcriptome profiling," *Appl. Environ. Microbiol.*, 2003, doi: 10.1128/AEM.69.8.4737-4742.2003.

[202] F. T. Haddadin and S. W. Harcum, "Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*," *Biotechnol. Bioeng.*, 2005, doi: 10.1002/bit.20340.

[203] J. Bonomo and R. T. Gill, "Amino acid content of recombinant proteins influences the metabolic burden response," *Biotechnol. Bioeng.*, 2005, doi: 10.1002/bit.20436.

[204] S. W. Harcum and F. T. Haddadin, "Global transcriptome response of recombinant *Escherichia coli* to heat-shock and dual heat-shock

recombinant protein induction," *J. Ind. Microbiol. Biotechnol.*, 2006, doi: 10.1007/s10295-006-0122-3.

[205] H. E. Smith, "The transcriptional response of *Escherichia coli* to recombinant protein insolubility," *J. Struct. Funct. Genomics*, 2007, doi: 10.1007/s10969-007-9030-7.

[206] F. T. Haddadin, H. Kurtz, and S. W. Harcum, "Serine hydroxamate and the transcriptome of high cell density recombinant *escherichia coli* MG1655," *Appl. Biochem. Biotechnol.*, 2009, doi: 10.1007/s12010-008-8241-0.

[207] J. H. Baek, M. J. Han, S. H. Lee, and S. Y. Lee, "Enhanced display of lipase on the *Escherichia coli* Cell surface, based on transcriptome analysis," *Appl. Environ. Microbiol.*, 2010, doi: 10.1128/AEM.02463-09.

[208] A. K. Sharma, S. Mahalik, C. Ghosh, A. B. Singh, and K. J. Mukherjee, "Comparative transcriptomic profile analysis of fed-batch cultures expressing different recombinant proteins in *Escherichia coli*," *AMB Express*, 2011, doi: 10.1186/2191-0855-1-33.

[209] F. Gubellini *et al.*, "Physiological response to membrane protein overexpression in *E. coli*," *Mol. Cell. Proteomics*, 2011, doi: 10.1074/mcp.M111.007930.

[210] S. Singh and A. J. Darwin, "FtsH-dependent degradation of phage shock protein C in Yersinia enterocolitica and *Escherichia coli*," *J. Bacteriol.*, 2011, doi: 10.1128/JB.05942-11.

[211] J. Tan *et al.*, "Independent component analysis of *E. coli*'s transcriptome reveals the cellular processes that respond to heterologous gene expression," *Metab. Eng.*, 2020, doi: 10.1016/j.ymben.2020.07.002.

[212] U. Rinas, "Synthesis rates of cellular proteins involved in translation and protein folding are strongly altered in response to overproduction of basic fibroblast growth factor by recombinant *Escherichia coli*," *Biotechnol. Prog.*, 1996, doi: 10.1021/bp9600039.

[213] F. Hoffmann and U. Rinas, "On-line estimation of the metabolic burden resulting from the synthesis of plasmid-encoded and heat-shock proteins by monitoring respiratory energy generation," *Biotechnol. Bioeng.*, 2001, doi: 10.1002/bit.10098.

[214] J. Weber, F. Hoffmann, and U. Rinas, "Metabolic adaptation of *Escherichia coli* during temperature-induced recombinant protein production: 2. Redirection of metabolic fluxes," *Biotechnol. Bioeng.*, 2002, doi: 10.1002/bit.10380.

[215] K. M. Champion *et al.*, "Comparison of the *Escherichia coli* proteomes for recombinant human growth hormone producing and nonproducing fermentations," *Proteomics*, vol. 3, no. 7, pp. 1365–1373, 2003, doi: 10.1002/pmic.200300430.

[216] M. J. Han, K. J. Jeong, J. S. Yoo, and S. Y. Lee, "Engineering *Escherichia coli* for increased productivity of serine-rich proteins based on proteome profiling," *Appl. Environ. Microbiol.*, vol. 69, no. 10, pp. 5772–5781, 2003, doi: 10.1128/aem.69.10.5772-5781.2003.

[217] K. Dürrschmid *et al.*, "Monitoring of protein profiles for the optimization of recombinant fermentation processes using public domain databases," *Electrophoresis*, 2003, doi: 10.1002/elps.200390027.

[218] D. H. Lee, S. G. Kim, Y. C. Park, S. W. Nam, K. H. Lee, and J. H. Seo, "Proteome analysis of recombinant *Escherichia coli* producing human glucagon-like peptide-1," *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, 2007, doi: 10.1016/j.jchromb.2006.09.042.

[219] S. Wagner *et al.*, "Consequences of membrane protein overexpression in *Escherichia coli*," *Mol. Cell. Proteomics*, 2007, doi: 10.1074/mcp.M600431-MCP200.

[220] S. Wagner *et al.*, "Tuning *Escherichia coli* for membrane protein overexpression," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 38, pp. 14371–14376, 2008, doi: 10.1073/pnas.0804090105.

[221] C. H. Cheng and W. C. Lee, "Protein solubility and differential proteomic profiling of recombinant *Escherichia coli* overexpressing double-tagged fusion proteins," *Microb. Cell Fact.*, 2010, doi: 10.1186/1475-2859-9-63.

[222] L. Ping *et al.*, "Quantitative proteomics reveals significant changes in cell shape and an energy shift after IPTG induction via an optimized SILAC approach for *Escherichia coli*," *J. Proteome Res.*, 2013, doi: 10.1021/pr400775w.

[223] Z. Li, M. Nimtz, and U. Rinas, "Global proteome response of *Escherichia coli* BL21 to production of human basic fibroblast growth factor in complex and defined medium," *Eng. Life Sci.*, 2017, doi: 10.1002/elsc.201700036.

[224] C. Wittmann, J. Weber, E. Betiku, J. Krömer, D. Böhm, and U. Rinas, "Response of fluxome and metabolome to temperature-induced recombinant protein synthesis in *Escherichia coli*," *J. Biotechnol.*, 2007, doi: 10.1016/j.jbiotec.2007.07.495.

[225] S. Carneiro, S. G. Villas-Bôas, E. C. Ferreira, and I. Rocha, "Metabolic footprint analysis of recombinant *Escherichia coli* strains during fed-batch fermentations," *Mol. Biosyst.*, 2011, doi: 10.1039/c0mb00143k.

[226] J. Heyland, L. M. Blank, and A. Schmid, "Quantification of metabolic limitations during recombinant protein production in *Escherichia coli*," *J. Biotechnol.*, 2011, doi: 10.1016/j.jbiotec.2011.06.016.

[227] H. Muhamadali *et al.*, "Metabolomic analysis of riboswitch containing *E. coli* recombinant expression system," *Mol. Biosyst.*, 2016, doi: 10.1039/c5mb00624d.

[228] Y. K. Chae, S. H. Kim, and J. L. Markley, "Relationship between recombinant protein expression and host metabolome as determined by two-dimensional NMR spectroscopy," *PLoS One*, 2017, doi: 10.1371/journal.pone.0177233.

[229] B. Jürgen *et al.*, "Monitoring of genes that respond to overproduction of an insoluble recombinant protein in *Escherichia coli* glucose-limited fed-batch fermentations," *Biotechnol. Bioeng.*, 2000, doi: 10.1002/1097-0290(20001020)70:2<217::AID-BIT11>3.0.CO;2-W.

[230] K. Dürrschmid *et al.*, "Monitoring of transcriptome and proteome profiles to investigate the cellular response of *E. coli* towards recombinant protein expression under defined chemostat conditions," *J. Biotechnol.*, 2008, doi: 10.1016/j.jbiotec.2008.02.013.

[231] S. Mahalik, A. K. Sharma, P. Jain, and K. J. Mukherjee, "Identifying genomic targets for protein over-expression by 'omics' analysis of Quiescent

Escherichia coli cultures," *Microb. Cell Fact.*, 2017, doi: 10.1186/s12934-017-0744-3.

[232] Z. Li *et al.*, "Transcriptional network analysis identifies key elements governing the recombinant protein production provoked reprogramming of carbon and energy metabolism in *Escherichia coli* BL21 (DE3)," *Eng. Reports*, vol. 3, no. 9, p. e12393, Sep. 2021, doi: 10.1002/ENG2.12393.

[233] F. Hoffmann, J. Weber, and U. Rinas, "Metabolic adaptation of *Escherichia coli* during tempe1. Champion KM, Nishihara JC, Joly JC, Arnott D. Similarity of the *Escherichia coli* proteome upon completion of different biopharmaceutical fermentation processes. Proteomics. 2001;1(9):1133–48. rature-," *Biotechnol. Bioeng.*, 2002, doi: 10.1002/bit.10379.

[234] A. B. Singh, A. K. Sharma, and K. J. Mukherjee, "Analyzing the metabolic stress response of recombinant *Escherichia coli* cultures expressing human interferon-beta in high cell density fed batch cultures using time course transcriptomic data," *Mol. Biosyst.*, 2012, doi: 10.1039/c1mb05414g.

[235] S. H. Yoon, M. J. Han, S. Y. Lee, K. J. Jeong, and J. S. Yoo, "Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture," *Biotechnol. Bioeng.*, vol. 81, no. 7, pp. 753–767, 2003, doi: 10.1002/bit.10626.

[236] S. H. Yoon, M. J. Han, S. Y. Lee, K. J. Jeong, and J. S. Yoo, "Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture," *Biotechnol. Bioeng.*, 2003, doi: 10.1002/bit.10626.

[237] K. A. Datsenko and B. L. Wanner, "One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products," *Proc. Natl. Acad. Sci. U. S. A.*, 2000, doi: 10.1073/pnas.120163297.

[238] "Promoters/Catalog/Anderson - parts.igem.org." http://parts.igem.org/Promoters/Catalog/Anderson (accessed Aug. 24, 2022).

[239] K. A. Datsenko and B. L. Wanner, "One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 12, pp. 6640–6645, 2000, doi: 10.1073/pnas.120163297.

[240] S. Andrews, "Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data," 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed Aug. 16, 2022).

[241] P. Ewels, M. Magnusson, S. Lundin, and M. Käller, "MultiQC: summarize analysis results for multiple tools and samples in a single report," *Bioinformatics*, vol. 32, no. 19, pp. 3047–3048, Oct. 2016, doi: 10.1093/BIOINFORMATICS/BTW354.

[242] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: A flexible trimmer for Illumina sequence data," *Bioinformatics*, 2014, doi: 10.1093/bioinformatics/btu170.

[243] K. L. Howe *et al.*, "Ensembl Genomes 2020-enabling non-vertebrate genomic research," *Nucleic Acids Res.*, 2020, doi: 10.1093/nar/gkz890.

[244] A. Dobin *et al.*, "STAR: Ultrafast universal RNA-seq aligner," *Bioinformatics*, 2013, doi: 10.1093/bioinformatics/bts635.

[245] Y. Liao, G. K. Smyth, and W. Shi, "The R package Rsubread is easier, faster,

cheaper and better for alignment and quantification of RNA sequencing reads," *Nucleic Acids Res.*, 2019, doi: 10.1093/nar/gkz114.

[246] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol.*, 2014, doi: 10.1186/s13059-014-0550-8.

[247] G. R. Warnes *et al.*, "gplots: Various R Programming Tools for Plotting Data. R package version 3.1.3." 2022, [Online]. Available: https://cran.r-project.org/package=gplots.

[248] K. Blighe, S. Rana, and M. Lewis, "EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.8.0.," 2020. https://github.com/kevinblighe/EnhancedVolcano (accessed Aug. 24, 2022).

[249] S. Tyanova, T. Temu, and J. Cox, "The MaxQuant computational platform for mass spectrometry-based shotgun proteomics," *Nat. Protoc. 2016 1112*, vol. 11, no. 12, pp. 2301–2319, Oct. 2016, doi: 10.1038/nprot.2016.136.

[250] E. Theodorakis *et al.*, "ProteoSign v2: a faster and evolved user-friendly online tool for statistical analyses of differential proteomics," *Nucleic Acids Res.*, vol. 49, no. W1, pp. W573–W577, Jul. 2021, doi: 10.1093/NAR/GKAB329.

[251] I. M. Keseler *et al.*, "The EcoCyc Database in 2021," *Front. Microbiol.*, vol. 12, p. 711077, Jul. 2021, doi: 10.3389/FMICB.2021.711077.

[252] S. Paley, K. Parker, A. Spaulding, J. F. Tomb, P. O'Maille, and P. D. Karp, "The omics dashboard for interactive exploration of gene-expression data," *Nucleic Acids Res.*, 2017, doi: 10.1093/nar/gkx910.

[253] P. D. Karp, "The EcoCyc Database," *EcoSal Plus*, vol. 6, no. 1, 2014.

[254] H. M. Shahzad Asif, M. D. Rolfe, J. Green, N. D. Lawrence, M. Rattray, and G. Sanguinetti, "TFInfer: A tool for probabilistic inference of transcription factor activities," *Bioinformatics*, 2010, doi: 10.1093/bioinformatics/btq469.

[255] T. Wu *et al.*, "clusterProfiler 4.0: A universal enrichment tool for interpreting omics data," *Innov.*, vol. 2, no. 3, p. 100141, Aug. 2021, doi: 10.1016/J.XINN.2021.100141.

[256] H. Thie, T. Schirrmann, M. Paschke, S. Dübel, and M. Hust, "SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in *E. coli*," *N. Biotechnol.*, vol. 25, no. 1, pp. 49–54, Jun. 2008, doi: 10.1016/J.NBT.2008.01.001.

[257] D. W. Selinger, R. M. Saxena, K. J. Cheung, G. M. Church, and C. Rosenow, "Global RNA Half-Life Analysis in *Escherichia coli* Reveals Positional Patterns of Transcript Degradation," *Genome Res.*, vol. 13, no. 2, p. 216, Feb. 2003, doi: 10.1101/GR.912603.

[258] M. Miot and J. M. Betton, "Protein quality control in the bacterial periplasm," *Microbial Cell Factories*. 2004, doi: 10.1186/1475-2859-3-4.

[259] Y. Zhou *et al.*, "Enhancing full-length antibody production by signal peptide engineering," *Microb. Cell Fact.*, vol. 15, no. 1, pp. 1–11, Mar. 2016, doi: 10.1186/S12934-016-0445-3/FIGURES/5.

[260] M. B. Stead *et al.*, "RNAsnap™: A rapid, quantitative and inexpensive, method for isolating total RNA from bacteria," *Nucleic Acids Res.*, 2012, doi:

10.1093/nar/gks680.

[261] A. P. Diz, M. Truebano, and D. O. F. Skibinski, “The consequences of sample pooling in proteomics: an empirical study,” *Electrophoresis*, vol. 30, no. 17, pp. 2967–2975, Sep. 2009, doi: 10.1002/ELPS.200900210.

[262] T. Dau, G. Bartolomucci, and J. Rappsilber, “Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin,” *Anal. Chem.*, vol. 92, no. 14, pp. 9523–9527, Jul. 2020, doi: 10.1021/ACS.ANALCHEM.0C00478/ASSET/IMAGES/LARGE/AC0C00478_0004.JPEG.

[263] J. P. Webb *et al.*, “Multi-omic based production strain improvement (MOBpsi) for bio-manufacturing of toxic chemicals,” *Metab. Eng.*, vol. 72, pp. 133–149, Jul. 2022, doi: 10.1016/J.YMBEN.2022.03.004.

[264] S. Mubeen, A. Tom Kodamullil, M. Hofmann-Apitius, and D. Domingo-Fernández, “On the influence of several factors on pathway enrichment analysis,” *Brief. Bioinform.*, vol. 23, no. 3, pp. 1–13, May 2022, doi: 10.1093/BIB/BBAC143.

[265] G. Orfanoudaki and A. Economou, “Proteome-wide subcellular topologies of *E. coli* polypeptides database (STEPdb),” *Mol. Cell. Proteomics*, vol. 13, no. 12, pp. 3674–3687, Dec. 2014, doi: 10.1074/MCP.O114.041137.

[266] A. Santos-Zavaleta *et al.*, “RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D212–D220, Jan. 2019, doi: 10.1093/NAR/GKY1077.

[267] T. L. Raivio and T. J. Silhavy, “The σ(E) and Cpx regulatory pathways: Overlapping but distinct envelope stress responses,” *Curr. Opin. Microbiol.*, 1999, doi: 10.1016/S1369-5274(99)80028-9.

[268] T. L. Raivio and T. J. Silhavy, “Periplasmic stress and ECF sigma factors,” *Annual Review of Microbiology*. 2001, doi: 10.1146/annurev.micro.55.1.591.

[269] T. L. Raivio, “Everything old is new again: An update on current research on the Cpx envelope stress response,” *Biochim. Biophys. Acta - Mol. Cell Res.*, 2014, doi: 10.1016/j.bbamcr.2013.10.018.

[270] M. Grabowicz and T. J. Silhavy, “Envelope Stress Responses: An Interconnected Safety Net,” *Trends in Biochemical Sciences*. 2017, doi: 10.1016/j.tibs.2016.10.002.

[271] T. L. Raivio and T. J. Silhavy, “Periplasmic stress and ECF sigma factors,” *Annu. Rev. Microbiol.*, vol. 55, pp. 591–624, 2001, doi: 10.1146/annurev.micro.55.1.591.

[272] C. Beloin *et al.*, “Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression,” *Mol. Microbiol.*, vol. 51, no. 3, pp. 659–674, Feb. 2004, doi: 10.1046/J.1365-2958.2003.03865.X.

[273] D. S. Ow, D. Y. Lim, P. Morin Nissom, A. Camattari, and V. V. Wong, “Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production,” *Microb. Cell Fact.*, vol. 9, no. 1, pp. 1–14, Apr. 2010, doi: 10.1186/1475-2859-9-22/TABLES/3.

[274] R. L. Guest, E. A. Court, J. L. Waldon, K. A. Schock, and T. L. Raivio, “Impaired

Efflux of the Siderophore Enterobactin Induces Envelope Stress in *Escherichia coli*," *Front. Microbiol.*, 2019, doi: 10.3389/fmicb.2019.02776.

[275] W. Luo and C. Brouwer, "Pathview: an R/Bioconductor package for pathway-based data integration and visualization," *Bioinformatics*, vol. 29, no. 14, pp. 1830–1831, Jul. 2013, doi: 10.1093/BIOINFORMATICS/BTT285.

[276] W. Luo, G. Pant, Y. K. Bhavnasi, S. G. Blanchard, and C. Brouwer, "Pathview Web: user friendly pathway visualization and data integration," *Nucleic Acids Res.*, vol. 45, no. W1, pp. W501–W508, Jul. 2017, doi: 10.1093/NAR/GKX372.

[277] H. Y. Lin *et al.*, "Cellular responses to strong overexpression of recombinant genes in *escherichia coli* - DNA relaxation and cell death after induction of alpha-glucosidase," *Recomb. Protein Prod. with Prokaryotic Eukaryot. Cells a Comp. View Host Physiol.*, pp. 55–73, 2001.

[278] D. Kornitzer, D. Teff, S. Altuvia, and A. B. Oppenheim, "Isolation, characterization, and sequence of an *Escherichia coli* heat shock gene, htpX.," *J. Bacteriol.*, vol. 173, no. 9, p. 2944, 1991, doi: 10.1128/JB.173.9.2944-2953.1991.

[279] M. Sakoh, K. Ito, and Y. Akiyama, "Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from *Escherichia coli*," *J. Biol. Chem.*, vol. 280, no. 39, pp. 33305–33310, Sep. 2005, doi: 10.1074/JBC.M506180200.

[280] J. Van Stelten, F. Silva, D. Belin, and T. J. Silhavy, "Effects of antibiotics and a proto-oncogene homolog on destruction of protein translocator SecY," *Science (80-.).*, 2009, doi: 10.1126/science.1172221.

[281] A. Kihara, Y. Akiyama, and K. Ito, "Different pathways for protein degradation by the FtsH/HflKC membrane-embedded protease complex: an implication from the interference by a mutant form of a new substrate protein, YccA," *J. Mol. Biol.*, vol. 279, no. 1, pp. 175–188, May 1998, doi: 10.1006/JMBI.1998.1781.

[282] S. L. Vogt and T. L. Raivio, "Hfq reduces envelope stress by controlling expression of envelope-localized proteins and protein complexes in enteropathogenic *Escherichia coli*," *Mol. Microbiol.*, vol. 92, no. 4, pp. 681–697, 2014, doi: 10.1111/MMI.12581.

[283] R. L. Guest, J. Wang, J. L. Wong, and T. L. Raivio, "A bacterial stress response regulates respiratory protein complexes to control envelope stress adaptation," *J. Bacteriol.*, 2017, doi: 10.1128/JB.00153-17.

[284] S. Schlegel, E. Rujas, A. J. Ytterberg, R. A. Zubarev, J. Luirink, and J. W. de Gier, "Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels," *Microb. Cell Fact.*, vol. 12, 2013, doi: 10.1186/1475-2859-12-24.

[285] R. Guleria, P. Jain, M. Verma, and K. J. Mukherjee, "Designing next generation recombinant protein expression platforms by modulating the cellular stress response in *Escherichia coli*," *Microb. Cell Fact.*, vol. 19, no. 1, pp. 1–17, Dec. 2020, doi: 10.1186/S12934-020-01488-W/TABLES/1.

[286] A. K. Sharma, J. Phue, E. Khatipov, N. Dalal, E. D. Anderson, and J. Shiloach, "Effect of restricted dissolved oxygen on expression of Clostridium difficile

toxin A subunit from *E. coli*," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/S41598-020-59978-1.

[287] J. Huangfu *et al.*, "Omics Analysis Reveals the Mechanism of Enhanced Recombinant Protein Production Under Simulated Microgravity," *Front. Bioeng. Biotechnol.*, vol. 8, p. 30, Feb. 2020, doi: 10.3389/FBIOE.2020.00030/BIBTEX.

[288] S. Ghatak, Z. A. King, A. Sastry, and B. O. Palsson, "The y-ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function," *Nucleic Acids Res.*, 2019, doi: 10.1093/nar/gkz030.

[289] A. Sivashanmugam, V. Murray, C. Cui, Y. Zhang, J. Wang, and Q. Li, "Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*," *Protein Sci.*, 2009, doi: 10.1002/pro.102.

[290] G. Sezonov, D. Joseleau-Petit, and R. D'Ari, "*Escherichia coli* physiology in Luria-Bertani broth," *J. Bacteriol.*, vol. 189, no. 23, pp. 8746–8749, 2007, doi: 10.1128/jb.01368-07.

[291] A. J. Darwin, "The phage-shock-protein response," *Mol. Microbiol.*, vol. 57, no. 3, pp. 621–628, Aug. 2005, doi: 10.1111/J.1365-2958.2005.04694.X.

[292] N. Joly *et al.*, "Managing membrane stress: The phage shock protein (Psp) response, from molecular mechanisms to physiology," *FEMS Microbiology Reviews*. 2010, doi: 10.1111/j.1574-6976.2010.00240.x.

[293] N. Shimohata, S. Chiba, N. Saikawa, K. Ito, and Y. Akiyama, "The Cpx stress response system of *Escherichia coli* senses plasma membrane proteins and controls HtpX, a membrane protease with a cytosolic active site," *Genes to Cells*, 2002, doi: 10.1046/j.1365-2443.2002.00554.x.

[294] T. L. Raivio, D. L. Popkin, and T. J. Silhavy, "The Cpx envelope stress response is controlled by amplification and feedback inhibition," *J. Bacteriol.*, vol. 181, no. 17, pp. 5263–5272, 1999, doi: 10.1128/JB.181.17.5263-5272.1999.

[295] C. V. Goemans, D. Vertommen, R. Agrebi, and J. F. Collet, "CnoX Is a Chaperedoxin: A Holdase that Protects Its Substrates from Irreversible Oxidation," *Mol. Cell*, vol. 70, no. 4, pp. 614-627.e7, May 2018, doi: 10.1016/J.MOLCEL.2018.04.002.

[296] R. E. Dalbey, P. Wang, and J. M. van Dijl, "Membrane proteases in the bacterial protein secretion and quality control pathway," *Microbiol. Mol. Biol. Rev.*, vol. 76, no. 2, pp. 311–330, Jun. 2012, doi: 10.1128/MMBR.05019-11.

[297] T. Baba *et al.*, "Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection," *Mol. Syst. Biol.*, 2006, doi: 10.1038/msb4100050.

[298] E. C. A. Goodall *et al.*, "The Essential Genome of *Escherichia coli* K-12," *MBio*, vol. 9, no. 1, Jan. 2018, doi: 10.1128/MBIO.02096-17.

[299] Y. Akiyama, A. Kihara, H. Mori, T. Ogura, and K. Ito, "Roles of the Periplasmic Domain of *Escherichia coli*FtsH (HflB) in Protein Interactions and Activity Modulation," *J. Biol. Chem.*, vol. 273, no. 35, pp. 22326–22333, Aug. 1998, doi: 10.1074/JBC.273.35.22326.

[300] A. Kihara, Y. Akiyama, and K. Ito, "A protease complex in the *Escherichia coli* plasma membrane: HflKC (HflA) forms a complex with FtsH (HflB),

regulating its proteolytic activity against SecY.," *EMBO J.*, 1996, doi: 10.1002/j.1460-2075.1996.tb01000.x.

[301] J. Domka, J. Lee, and T. K. Wood, "YliH (BssR) and YceP (BssS) regulate *Escherichia coli* K-12 biofilm formation by influencing cell signaling," *Appl. Environ. Microbiol.*, vol. 72, no. 4, pp. 2449–2459, Apr. 2006, doi: 10.1128/AEM.72.4.2449-2459.2006.

[302] C. Beloin *et al.*, "Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression," *Mol. Microbiol.*, vol. 51, no. 3, pp. 659–674, Feb. 2004, doi: 10.1046/J.1365-2958.2003.03865.X.

[303] D. Ren, L. A. Bedzyk, S. M. Thomas, R. W. Ye, and T. K. Wood, "Gene expression in *Escherichia coli* biofilms," *Appl. Microbiol. Biotechnol.*, vol. 64, no. 4, pp. 515–524, May 2004, doi: 10.1007/S00253-003-1517-Y.

[304] M. A. Schembri, K. Kjærgaard, and P. Klemm, "Global gene expression in *Escherichia coli* biofilms," *Mol. Microbiol.*, vol. 48, no. 1, pp. 253–267, Apr. 2003, doi: 10.1046/J.1365-2958.2003.03432.X.

[305] G. Jovanovic *et al.*, "The N-terminal amphipathic helices determine regulatory and effector functions of phage shock protein A (PspA) in *Escherichia coli*," *J. Mol. Biol.*, vol. 426, no. 7, pp. 1498–1511, Apr. 2014, doi: 10.1016/J.JMB.2013.12.016.

[306] Y. Kawano *et al.*, "Enhancement of L-cysteine production by disruption of yciW in *Escherichia coli*," *J. Biosci. Bioeng.*, vol. 119, no. 2, pp. 176–179, Feb. 2015, doi: 10.1016/J.JBIOSC.2014.07.006.

[307] S. Morigasaki, A. Umeyama, Y. Kawano, Y. Aizawa, and I. Ohtsu, "Defect of RNA pyrophosphohydrolase RppH enhances fermentative production of L-cysteine in *Escherichia coli*," *J. Gen. Appl. Microbiol.*, vol. 66, no. 6, pp. 307–314, 2021, doi: 10.2323/JGAM.2019.12.004.

[308] D. J. Lee *et al.*, "Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains," *Bmc Microbiol.*, vol. 9, 2009, doi: 10.1186/1471-2180-9-252.

[309] J. R. Kelly *et al.*, "Measuring the activity of BioBrick promoters using an in vivo reference standard," *J. Biol. Eng.*, vol. 3, no. 1, pp. 1–13, Apr. 2009, doi: 10.1186/1754-1611-3-4/FIGURES/3.

[310] M. Kitagawa *et al.*, "Complete set of ORF clones of *Escherichia coli* ASKA library (A Complete S et of *E. coli* K -12 ORF A rchive): Unique Resources for Biological Research," *DNA Res.*, vol. 12, no. 5, pp. 291–299, Jan. 2005, doi: 10.1093/DNARES/DSI012.

[311] B. R. Glick, "Metabolic load and heterologous gene expression," *Biotechnol. Adv.*, vol. 13, no. 2, pp. 247–261, Jan. 1995, doi: 10.1016/0734-9750(95)00004-A.

[312] A. J. Link, D. Phillips, and G. M. Church, "Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: Application to open reading frame characterization," *J. Bacteriol.*, 1997, doi: 10.1128/jb.179.20.6228-6237.1997.

[313] V. A. Karkhanis, A. P. Mascarenhas, and S. A. Martinis, "Amino Acid Toxicities of *Escherichia coli* That Are Prevented by Leucyl-tRNA Synthetase Amino Acid Editing," *J. Bacteriol.*, vol. 189, no. 23, p. 8765, Dec. 2007, doi: 10.1128/JB.01215-07.

[314] S. Korshunov, K. R. C. Imlay, and J. A. Imlay, "Cystine import is a valuable but risky process whose hazards *Escherichia coli* minimizes by inducing a cysteine exporter," *Mol. Microbiol.*, vol. 113, no. 1, pp. 22–39, Jan. 2020, doi: 10.1111/MMI.14403.

[315] C. Kari, Z. Nagy, and F. Hernádi, "Effect of cysteine on inducible synthesis of β-galactosidase in *Escherichia coli*," *Biochem. Pharmacol.*, vol. 20, no. 5, pp. 975–978, May 1971, doi: 10.1016/0006-2952(71)90319-4.

[316] S. M. Alonso Villela, H. Kraïem, B. Bouhaouala-Zahar, C. Bideaux, C. A. Aceves Lara, and L. Fillaudeau, "A protocol for recombinant protein quantification by densitometry," *Microbiologyopen*, vol. 9, no. 6, pp. 1175–1182, Jun. 2020, doi: 10.1002/MBO3.1027.

[317] P. Rehbein and H. Schwalbe, "Integrated protocol for reliable and fast quantification and documentation of electrophoresis gels," *Protein Expr. Purif.*, vol. 110, pp. 1–6, 2015, doi: 10.1016/J.PEP.2014.12.006.

[318] J. Van Stelten, F. Silva, D. Belin, and T. J. Silhavy, "Effects of antibiotics and a proto-oncogene homolog on destruction of protein translocator SecY," *Science*, vol. 325, no. 5941, pp. 753–756, 2009, doi: 10.1126/SCIENCE.1172221.

[319] M.-J. Han and S. Y. Lee, "The *Escherichia coli* Proteome: Past, Present, and Future Prospects," *Microbiol. Mol. Biol. Rev.*, 2006, doi: 10.1128/mmbr.00036-05.

[320] M. Dragosits, D. Nicklas, and I. Tagkopoulos, "A synthetic biology approach to self-regulatory recombinant protein production in *Escherichia coli*," *J. Biol. Eng.*, vol. 6, no. 1, 2012, doi: 10.1186/1754-1611-6-2.

[321] A. Gawin *et al.*, "Construction and characterization of broad-host-range reporter plasmid suitable for on-line analysis of bacterial host responses related to recombinant protein production," *Microb. Cell Fact.*, vol. 18, no. 1, pp. 1–14, May 2019, doi: 10.1186/S12934-019-1128-7/TABLES/2.

7 Appendix

Appendix 1. Sequences of model periplasmic recombinant proteins (scFv, Hel4, HGH and sfGFP). Signal sequences are highlighted in bold.

>OmpA_signal_peptide+scFv

ATGAAGAAAACCGCTATTGCTATCGCTGTTGCTCTGGCAGGTTTTGCTACTGTT
GCCCAGGCCGATATTCAGATGACCCAATCCCCGAGCTCTCTGTCCGCGTCTGTGGGCG
ATCGTGTGACGATCACGTGCCGTGCGTCTCAGTCTATCGATTCTTATCTGCACTGGTA
CCAACAGAAACCAGGTAAGGCTCCGAAACTGCTGATCTATAGCGCGAGCGAACTGCA
ATCTGGTGTGCCGTCTCGTTTCTCCGGCTCTGGTTCCGGTACTGACTTCACCCTGACC
ATTAGCAGCCTGCAGCCGGAGGATTTCGCTACCTACTACTGTCAGCAGGTTGTCTGGC
GCCCTTTCACTTTCGGTCAGGGCACGAAGGTGGAGATCAAGCGCTAA

>OmpA_signal_peptide+Hel4

ATGAAGAAGACGGCAATCGCAATCGCAGTCGCTTTGGCAGGTTTTGCGACGGTT
GCGCAGGCAGAAGTGCAACTGCTGGAGAGCGGTGGCGGTCTGGTCCAACCGGGCGGT
AGCTTGCGCCTGTCGTGCGCCGCCTCTGGTTCTCGTATCTCTGACGAAGATATGGGTT
GGGTGCGCCAAGCTCCGGGCAAAGGTCTGGAGTGGGTTAGCAGCATTTACGGCCCGT
CCGGCTCCACGTACTATGCGGATAGCGATAAGGGTCGTTTCACCATCAGCCGTGACAA
TTCCAAAAACACGCTGTACCTGCAGATGAATAGCCTGCGTGCAGAGGACACCGCCGTT
TACTACTGCGCGAGCGCGCTGGAGCCTCTGAGCGAGCCGCTGGGTTATTGGGGTCAAG
GCACCTTGGTACACGGTTAG

>OmpA_signal_peptide+HGH

ATGAAAAAGACTGCTATTGCTATTGCGGTTGCTCTGGCTGGTTTCGCCACCGTGG
CGCAGGCATTCCCGACGATTCCGCTGTCTCGTCTGTTCGATAACGCTATGCTGCGTGC
ACATCGTCTGCACCAGCTGGCATTCGACACTTATCAAGAATTCGAGGAGGCCTACATC
CCGAAAGAACAGAAATACAGCTTCCTGCAGAACCCGCAAACCTCCCTGTGCTTCAGCG
AATCCATCCCGACTCCGTCTAACCGCGAAGAAACCCAGCAGAAATCTAATCTGGAACT
GCTGCGTATTAGCCTGCTGCTGATCCAGTCTTGGCTGGAACCGGTGCAATTCCTGCGC
AGCGTCTTCGCGAACTCCCTGGTATATGGCGCTTCCGACTCTAACGTATACGATCTGC
TGAAAGACCTGGAGGAGGGTATCCAGACCCTGATGGGCCGCCTGGAGGACGGCAGCC
CGCGTACTGGCCAGATCTTTAAACAGACTTACTCTAAATTCGACACCAACTCCCATAA
TGATGACGCCCTGCTGAAAAACTATGGTCTGCTGTACTGCTTCCGTAAAGACATGGA
CAAAGTTGAGACCTTCCTGCGTATTGTTCAGTGTCGCTCTGTTGAGGGTTCTTGCGGC
TTCCTGCAGGAAAACCTGTACTTCCAAGGTGACTACAAGGACGATGACGATAAGTGG
AGCCATCCGCAGTTTGAGAAATCTAGAAAGCTTTAA

>OmpA_signal_peptide+sfGFP+HIS_tag

ATGAAGAAGACGGCAATCGCAATCGCAGTCGCTTTGGCAGGTTTTGCGACGGTT
GCGCAGGCAGAAGTGCAACTGCTGGAGAGCGGTGGCGGTCTGGTCCAACCGGGCGGT
AGCTTGCGCCTGTCGTGCGCCGCCTCTGGTTTTCGTATCTCTGACGAAGATATGGGTT
GGGTGCGCCAAGCTCCGGGCAAAGGTCTGGAGTGGGTTAGCAGCATTTACGGCCCGT
CCGGCTCCACGTACTATGCGGATAGCGTTAAGGGTCGTTTCACCATCAGCCGTGACAA
TTCCAAAAACACGCTGTACCTGCAGATGAATAGCCTGCGTGCAGAGGACACCGCCGTT
TATTACTGCGCGAGCGCGCTGGAGCCACTGAGCGAGCCGCTGGGTTTTTGGGGTCAA
GGCACCTTGGTCACGGTTAGCTCGCTCGAGGCATTGTCCTCTTAG

Appendix 2. Sequence of synthesised genes. Restriction sites are underlined.

>HflCK_High

GAATTCGCGGCCGCTTCTAGATTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGCGTGGAATCAGCCCGGTAATAACGGACAAG
ACCGCGACCCGTGGGGAAGCAGCAAACCTGGCGGCAACTCTGAGGGAAATGGAAACA
AAGGCGGTCGCGATCAAGGGCCACCTGATTTAGATGATATCTTCCGCAAACTGAGCA
AAAAGCTCGGTGGTCTGGGCGGCGGTAAAGGCACCGGATCTGGCGGTGGCAGTTCAT
CGCAAGGCCCGCGCCCGCAGCTTGGCGGTCGTGTCGTTACCATCGCAGCGGCAGCGAT
TGTCATTATCTGGGCGGCCAGTGGTTTCTATACCATTAAAGAAGCCGAACGCGGCGT
GGTAACACGCTTTGGTAAATTCAGCCATCTGGTTGAGCCGGGTCTGAACTGGAAACC
GACGTTTATCGACGAAGTCAAACCGGTGAACGTGGAAGCCGTGCGTGAACTGGCCGC
TTCTGGTGTGATGCTGACGTCGGACGAGAACGTAGTGCGCGTTGAGATGAACGTGCA
GTACCGCGTCACCAATCCGGAAAAATATCTGTATAGCGTGACCAGCCCGGATGACAG
CCTGCGTCAGGCTACCGACAGCGCCCTGCGTGGAGTTATCGGTAAATACACCATGGAC
CGCATTCTGACGGAAGGTCGTACCGTGATTCGTAGCGATACTCAGCGCGAACTGGAA
GAGACGATTCGTCCGTATGACATGGGTATCACGCTGCTGGACGTCAACTTCCAGGCTG
CTCGTCCGCCGGAAGAAGTAAAAGCGGCGTTTGACGATGCGATTGCCGCGCGTGAAA
ACGAACAGCAATACATTCGTGAAGCAGAAGCGTATACCAACGAAGTTCAGCCGCGTG
CGAACGGTCAGGCGCAACGTATCCTCGAAGAGGCGCGTGCGTACAAGGCCCAGACCA
TCCTGGAAGCTCAGGGTGAAGTGGCGCGCTTTGCTAAACTTCTGCCGGAATATAAAG
CCGCGCCGGAAATTACTCGCGAGCGTCTGTATATCGAGACGATGGAAAAAGTGTTGG
GTAACACCCGCAAAGTGCTGGTTAACGATAAAGGTGGCAACCTGATGGTTCTGCCGT
TAGACCAGATGCTGAAAGGTGGTAACGCCCCTGCGGCGAAGAGCGATAACGGTGCCA
GCAATCTGCTGCGTCTGCCGCCAGCCTCTTCCTCCACAACCAGTGGAGCAAGCAACAC
GTCGTCCACCAGTCAGGGCGATATTATGGACCAACGCCGCGCCAACGCGCAGCGTAAC
GACTACCAGCGTCAGGGGGAATAACGATGCGTAAGTCAGTTATCGCGATTATCATCA
TCGTGCTGGTAGTGCTTTACATGTCTGTCTTTGTCGTCAAAGAAGGTGAGCGCGGTA
TTACGCTGCGTTTTGGTAAGGTACTGCGTGACGATGACAACAAACCTCTGGTTTATG
AGCCGGGTCTGCATTTCAAGATACCGTTCATTGAAACGGTGAAAATGCTCGACGCAC
GTATTCAGACCATGGACAACCAGGCCGACCGCTTTGTGACCAAAGAGAAGAAAGACC
TGATCGTCGACTCTTACATCAAATGGCGCATCAGCGATTTCAGCCGTTACTACCTGGC
AACGGGTGGTGGCGACATTTCGCAAGCGGAAGTGCTGTTGAAACGTAAGTTCTCTGA
CCGTCTGCGTTCTGAAATTGGTCGCCTGGACGTGAAAGATATCGTCACCGATTCCCGT
GGTCGTCTGACCCTCGAAGTACGTGACGCGCTGAACTCCGGTTCTGCGGGTACAGAAG
ATGAAGTTACTACCCCGGCGGCAGATAACGCCATTGCCGAAGCGGCAGAGCGCGTAA
CGGCTGAGACGAAGGGCAAAGTTCCGGTCATCAACCCGAACAGTATGGCGGCGCTGG
GTATTGAAGTTGTCGATGTGCGTATCAAGCAGATCAACCTGCCGACCGAAGTGTCTG
AAGCGATCTACAACCGTATGCGCGCCGAGCGTGAAGCGGTAGCGCGTCGTCACCGTTC
ACAAGGTCAGGAAGAAGCGGAAAAACTGCGCGCGACTGCCGACTATGAAGTGACCAG
AACGCTGGCAGAAGCTGAGCGTCAGGGCCGCATCATGCGTGGTGAAGGCGATGCCGA
AGCAGCCAAACTGTTTGCTGATGCATTCAGTAAAGATCCGGACTTCTACGCATTCAT
CCGTAGCCTGCGTGCTTATGAGAACAGCTTCTCTGGCAATCAGGACGTGATGGTCAT
GAGCCCGGATAGCGATTTCTTCCGCTACATGAAGACGCCGACTTCCGCAACGCGTTAA
CTGCAG

>HflCK_Medium

GAATTCGCGGCCGCTTCTAGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGCGTGGAATCAGCCCGGTAATAACGGACAAG
ACCGCGACCCGTGGGGAAGCAGCAAACCTGGCGGCAACTCTGAGGGAAATGGAAACA

AAGGCGGTCGCGATCAAGGGCCACCTGATTTAGATGATATCTTCCGCAAACTGAGCA
AAAAGCTCGGTGGTCTGGGCGGCGGTAAAGGCACCGGATCTGGCGGTGGCAGTTCAT
CGCAAGGCCCGCGCCCGCAGCTTGGCGGTCGTGTCGTTACCATCGCAGCGGCAGCGAT
TGTCATTATCTGGGCGGCCAGTGGTTTCTATACCATTAAAGAAGCCGAACGCGGCGT
GGTAACACGCTTTGGTAAATTCAGCCATCTGGTTGAGCCGGGTCTGAACTGGAAACC
GACGTTTATCGACGAAGTCAAACCGGTGAACGTGGAAGCCGTGCGTGAACTGGCCGC
TTCTGGTGTGATGCTGACGTCGGACGAGAACGTAGTGCGCGTTGAGATGAACGTGCA
GTACCGCGTCACCAATCCGGAAAAATATCTGTATAGCGTGACCAGCCCGGATGACAG
CCTGCGTCAGGCTACCGACAGCGCCCTGCGTGGAGTTATCGGTAAATACACCATGGAC
CGCATTCTGACGGAAGGTCGTACCGTGATTCGTAGCGATACTCAGCGCGAACTGGAA
GAGACGATTCGTCCGTATGACATGGGTATCACGCTGCTGGACGTCAACTTCCAGGCTG
CTCGTCCGCCGGAAGAAGTAAAAGCGGCGTTTGACGATGCGATTGCCGCGCGTGAAA
ACGAACAGCAATACATTCGTGAAGCAGAAGCGTATACCAACGAAGTTCAGCCGCGTG
CGAACGGTCAGGCGCAACGTATCCTCGAAGAGGCGCGTGCGTACAAGGCCCAGACCA
TCCTGGAAGCTCAGGGTGAAGTGGCGCGCTTTGCTAAACTTCTGCCGGAATATAAAG
CCGCGCCGGAAATTACTCGCGAGCGTCTGTATATCGAGACGATGGAAAAAGTGTTGG
GTAACACCCGCAAAGTGCTGGTTAACGATAAAGGTGGCAACCTGATGGTTCTGCCGT
TAGACCAGATGCTGAAAGGTGGTAACGCCCCTGCGGCGAAGAGCGATAACGGTGCCA
GCAATCTGCTGCGTCTGCCGCCAGCCTCTTCCTCCACAACCAGTGGAGCAAGCAACAC
GTCGTCCACCAGTCAGGGCGATATTATGGACCAACGCCGCGCCAACGCGCAGCGTAAC
GACTACCAGCGTCAGGGGGAATAACGATGCGTAAGTCAGTTATCGCGATTATCATCA
TCGTGCTGGTAGTGCTTTACATGTCTGTCTTTGTCGTCAAAGAAGGTGAGCGCGGTA
TTACGCTGCGTTTTGGTAAGGTACTGCGTGACGATGACAACAAACCTCTGGTTTATG
AGCCGGGTCTGCATTTCAAGATACCGTTCATTGAAACGGTGAAAATGCTCGACGCAC
GTATTCAGACCATGGACAACCAGGCCGACCGCTTTGTGACCAAAGAGAAGAAAGACC
TGATCGTCGACTCTTACATCAAATGGCGCATCAGCGATTTCAGCCGTTACTACCTGGC
AACGGGTGGTGGCGACATTTCGCAAGCGGAAGTGCTGTTGAAACGTAAGTTCTCTGA
CCGTCTGCGTTCTGAAATTGGTCGCCTGGACGTGAAAGATATCGTCACCGATTCCCGT
GGTCGTCTGACCCTCGAAGTACGTGACGCGCTGAACTCCGGTTCTGCGGGTACAGAAG
ATGAAGTTACTACCCCGGCGGCAGATAACGCCATTGCCGAAGCGGCAGAGCGCGTAA
CGGCTGAGACGAAGGGCAAAGTTCCGGTCATCAACCCGAACAGTATGGCGGCGCTGG
GTATTGAAGTTGTCGATGTGCGTATCAAGCAGATCAACCTGCCGACCGAAGTGTCTG
AAGCGATCTACAACCGTATGCGCGCCGAGCGTGAAGCGGTAGCGCGTCGTCACCGTTC
ACAAGGTCAGGAAGAAGCGGAAAAACTGCGCGCGACTGCCGACTATGAAGTGACCAG
AACGCTGGCAGAAGCTGAGCGTCAGGGCCGCATCATGCGTGGTGAAGGCGATGCCGA
AGCAGCCAAACTGTTTGCTGATGCATTCAGTAAAGATCCGGACTTCTACGCATTCAT
CCGTAGCCTGCGTGCTTATGAGAACAGCTTCTCTGGCAATCAGGACGTGATGGTCAT
GAGCCCGGATAGCGATTTCTTCCGCTACATGAAGACGCCGACTTCCGCAACGCGTTAA
CTGCAG

>YccA high

GAATTCGCGGCCGCTTCTAGATTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGATCGTATTGTTAGTTCTTCACATGACCGTA
CATCACTGCTTAGCACCCATAAGGTGCTGCGTAATACCTATTTTCTGCTGAGCCTGAC
GCTGGCCTTTTCGGCGATTACCGCAACTGCCAGTACGGTGCTGATGCTGCCATCTCCG
GGTCTGATTCTGACGCTGGTGGGTATGTATGGTTTGATGTTCCTGACCTATAAAACG
GCGAATAAGCCGACCGGGATTATCTCCGCATTCGCCTTTACCGGTTTTCTGGGTTATA
TCCTCGGACCTATTCTGAACACCTATCTGTCTGCCGGAATGGGTGACGTAATCGCTAT
GGCACTGGGCGGAACGGCGTTAGTGTTCTTCTGCTGCTCTGCATATGTGCTGACCACC
CGCAAAGATATGTCGTTCCTCGGCGGTATGCTGATGGCGGGTATTGTGGTGGTGCTG

ATTGGTATGGTTGCGAATATCTTCCTGCAGCTGCCTGCTCTGCATCTGGCGATCAGCG
CGGTCTTCATTCTGATCTCCTCTGGCGCTATCTTGTTTGAAACCAGCAACATCATTCA
TGGCGGTGAGACGAACTATATTCGTGCCACGGTTAGCCTGTATGTTTCGCTGTACAA
CATCTTCGTCAGCCTGCTGAGCATTCTGGGCTTCGCTAGCCGCGATTAACCCGGG

>YccA_medium

GAATTCGCGGCCGCTTCTAGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGATCGTATTGTTAGTTCTTCACATGACCGTA
CATCACTGCTTAGCACCCATAAGGTGCTGCGTAATACCTATTTTCTGCTGAGCCTGAC
GCTGGCCTTTTCGGCGATTACCGCAACTGCCAGTACGGTGCTGATGCTGCCATCTCCG
GGTCTGATTCTGACGCTGGTGGGTATGTATGGTTTGATGTTCCTGACCTATAAAACG
GCGAATAAGCCGACCGGGATTATCTCCGCATTCGCCTTTACCGGTTTTCTGGGTTATA
TCCTCGGACCTATTCTGAACACCTATCTGTCTGCCGGAATGGGTGACGTAATCGCTAT
GGCACTGGGCGGAACGGCGTTAGTGTTCTTCTGCTGCTCTGCATATGTGCTGACCACC
CGCAAAGATATGTCGTTCCTCGGCGGTATGCTGATGGCGGGTATTGTGGTGGTGCTG
ATTGGTATGGTTGCGAATATCTTCCTGCAGCTGCCTGCTCTGCATCTGGCGATCAGCG
CGGTCTTCATTCTGATCTCCTCTGGCGCTATCTTGTTTGAAACCAGCAACATCATTCA
TGGCGGTGAGACGAACTATATTCGTGCCACGGTTAGCCTGTATGTTTCGCTGTACAA
CATCTTCGTCAGCCTGCTGAGCATTCTGGGCTTCGCTAGCCGCGATTAACCCGGG

>BssS_high

GAATTCGCGGCCGCTTCTAGATTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGAAAAAAATAATGAAGTCATTCAGACTCAT
CCGCTCGTAGGGTGGGACATCAGCACCGTTGATAGCTATGATGCGCTGATGTTGCGT
TTGCACTACCAGACCCCAAATAAGTCCGAGCAGGAAGGGACTGAAGTTGGTCAGACG
CTCTGGTTAACCACTGATGTTGCCAGACAATTTATTTCGATATTAGAAGCAGGAATC
GCCAAAATTGAATCCGGTGATTTCCAGGTAAACGAGTATCGGCGTCATTAACTGCAG

>BssS_medium

GAATTCGCGGCCGCTTCTAGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGAAAAAAATAATGAAGTCATTCAGACTCAT
CCGCTCGTAGGGTGGGACATCAGCACCGTTGATAGCTATGATGCGCTGATGTTGCGT
TTGCACTACCAGACCCCAAATAAGTCCGAGCAGGAAGGGACTGAAGTTGGTCAGACG
CTCTGGTTAACCACTGATGTTGCCAGACAATTTATTTCGATATTAGAAGCAGGAATC
GCCAAAATTGAATCCGGTGATTTCCAGGTAAACGAGTATCGGCGTCATTAACTGCAG

>YjfJ_high

GAATTCGCGGCCGCTTCTAGATTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGGAATTTTAAAAAGTTTATTTACGCTGGGG
AAATCGTTTATCTCCCAGGCGGAAGAATCCATTGAAGAAACCCAGGGCGTGCGCATG
CTGGAACAGCATATTCGTGACGCTAAAGCTGAACTCGATAAAGCCGGAAAATCTCGC
GTTGATCTGCTGGCGCGGGTGAAATTAAGTCACGATAAGCTGAAAGATTTACGTGAG
CGCAAAGCCAGTCTGGAAGCCCGTGCGCTGGAAGCGTTAAGCAAGAACGTTAATCCG
TCGTTGATTAACGAAGTTGCTGAAGAAATCGCGCGCCTTGAGAATCTCATTACCGCT
GAAGAGCAAGTGTTGTCGAATCTGGAAGTCTCCCGTGATGGCGTGGAAAAAGCAGTT
ACAGCGACAGCGCAGCGTATCGCTCAGTTTGAGCAGCAAATGGAAGTCGTTAAAGCC
ACTGAAGCCATGCAGCGTGCACAACAGGCTGTAACAACCTCTACCGTTGGCGCATCTT
CCAGCGTTTCGACAGCGGCAGAATCCTTAAAACGCCTGCAAACGCGTCAGGCCGAACG
TCAGGCTCGCCTGGATGCTGCCGCACAGTTGGAGAAAGTCGCAGACGGTCGCGACCTT
GACGAAAAGCTGGCGGAAGCCGGAATTGGCGGTAGCAATAAAAGTAGCGCCCAGGAT
GTATTAGCAAGACTGCAACGCCAACAGGGCGAGTAACTGCAG

>YjfJ_medium

GAATTCGCGGCCGCTTCTAGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGGAATTTTAAAAAGTTTATTTACGCTGGGG
AAATCGTTTATCTCCCAGGCGGAAGAATCCATTGAAGAAACCCAGGGCGTGCGCATG
CTGGAACAGCATATTCGTGACGCTAAAGCTGAACTCGATAAAGCCGGAAAATCTCGC
GTTGATCTGCTGGCGCGGGTGAAATTAAGTCACGATAAGCTGAAAGATTTACGTGAG
CGCAAAGCCAGTCTGGAAGCCCGTGCGCTGGAAGCGTTAAGCAAGAACGTTAATCCG
TCGTTGATTAACGAAGTTGCTGAAGAAATCGCGCGCCTTGAGAATCTCATTACCGCT
GAAGAGCAAGTGTTGTCGAATCTGGAAGTCTCCCGTGATGGCGTGGAAAAAGCAGTT
ACAGCGACAGCGCAGCGTATCGCTCAGTTTGAGCAGCAAATGGAAGTCGTTAAAGCC
ACTGAAGCCATGCAGCGTGCACAACAGGCTGTAACAACCTCTACCGTTGGCGCATCTT
CCAGCGTTTCGACAGCGGCAGAATCCTTAAAACGCCTGCAAACGCGTCAGGCCGAACG
TCAGGCTCGCCTGGATGCTGCCGCACAGTTGGAGAAAGTCGCAGACGGTCGCGACCTT
GACGAAAAGCTGGCGGAAGCCGGAATTGGCGGTAGCAATAAAAGTAGCGCCCAGGAT
GTATTAGCAAGACTGCAACGCCAACAGGGCGAGTAACTGCAG

>YciW_high

GAATTCGCGGCCGCTTCTAGATTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGAACAACGCCACATCACCGGCAAAAGCCACT
GGTATCATGAAACGCAATCCAGTACTACGGAGTATGACGTTCTGCCTCTGGTCCCGG
AAGCCGCAAAGGTCAGCGATCCCTTTCTACTCGACGTGATCCTTGAAAAAGAAACGC
TGGCCCCCTTCCTTTCATGGCTGGACCCTGCGCGTGTTCTTGCAGTGGATTTGTTCCC
TGACCAGCTTACCGTGACCCGTTCACAGACCTTCACCGCTTATGAACGCTTGTCGACG
GCCCTGACGGTTGCTCAGGTTTGCGGCGTCCAGCGGTTATGTAACTACTATTCGGCGC
GACTTACGCCGCTCCCCGGGCCTGATTCCACCAGGGAAAGTAATCATCGGTTGGCACA
AATCACGCAATATGCCCGCCAACTGGCTAGCTCGCCTTCTATTATCGACAACCGATCG
CGCCAGCATCTGAATGACGTCGGTCTTACTGCCTGGGACTGTGTGATCATTAGCCAAA
TCATTGGTTTTATTGGCTTTCAGGCGCGGACAATTGCGACATTTCAGGCTTATCTCGG
GCATCCGGTACGCTGGTTACCCGGGCTGGAGATACAAAACTACGCCGACGCGTCACTG
TTTGCTGATGAATCATTACGCTGGCGAAGCAGCTATGAGGTGGAAAAACTACCTGAA
GAGCACACAAAAAGTTCAACTGCAGAACTTTGCCAACTGGCCGAAATACTCTCTCTCC
ACCCTATTTCACTTTCCCTTCTCGAAAAGTTGTTAAACAGCACACGGGGCAATACACA
GCCGGATAATCAGCTTGCGGCGTTGTTATGCGCGCGTATAAATGGCAGTCCTGCTTG
TTTTGCCACCTGTATGGATTCATCAAATGAATATAAAAAAATCAGCACCCTTATGCG
CAAGGGCGAAAATGAAATTAACCAATGGGCTGACCGTCATTCTGTTGAGCGCGCTAC
CGTTCAGGCGATACAATGGCTGACCCGAGCACCCGATCGCTTTAGCGCCGCCCAGTTC
AGCCCTTTACTCGAACACGAAAAATCATCAACGCAGATTATTAATCTGCTGGTATGG
AGCGGGCTGTGTGGCTGGATAAATCGCTTAAAAATCGCGTTGGGTGAGACATATTAA
CCCGGG

>YciW_medium

GAATTCGCGGCCGCTTCTAGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGGAACAACGCCACATCACCGGCAAAAGCCACT
GGTATCATGAAACGCAATCCAGTACTACGGAGTATGACGTTCTGCCTCTGGTCCCGG
AAGCCGCAAAGGTCAGCGATCCCTTTCTACTCGACGTGATCCTTGAAAAAGAAACGC
TGGCCCCCTTCCTTTCATGGCTGGACCCTGCGCGTGTTCTTGCAGTGGATTTGTTCCC
TGACCAGCTTACCGTGACCCGTTCACAGACCTTCACCGCTTATGAACGCTTGTCGACG
GCCCTGACGGTTGCTCAGGTTTGCGGCGTCCAGCGGTTATGTAACTACTATTCGGCGC
GACTTACGCCGCTCCCCGGGCCTGATTCCACCAGGGAAAGTAATCATCGGTTGGCACA
AATCACGCAATATGCCCGCCAACTGGCTAGCTCGCCTTCTATTATCGACAACCGATCG
CGCCAGCATCTGAATGACGTCGGTCTTACTGCCTGGGACTGTGTGATCATTAGCCAAA
TCATTGGTTTTATTGGCTTTCAGGCGCGGACAATTGCGACATTTCAGGCTTATCTCGG

GCATCCGGTACGCTGGTTACCCGGGCTGGAGATACAAAACTACGCCGACGCGTCACTG
TTTGCTGATGAATCATTACGCTGGCGAAGCAGCTATGAGGTGGAAAAACTACCTGAA
GAGCACACAAAAAGTTCAACTGCAGAACTTTGCCAACTGGCCGAAATACTCTCTCTCC
ACCCTATTTCACTTTCCCTTCTCGAAAAGTTGTTAAACAGCACACGGGGCAATACACA
GCCGGATAATCAGCTTGCGGCGTTGTTATGCGCGCGTATAAATGGCAGTCCTGCTTG
TTTTGCCACCTGTATGGATTCATCAAATGAATATAAAAAAATCAGCACCCTTATGCG
CAAGGGCGAAAATGAAATTAACCAATGGGCTGACCGTCATTCTGTTGAGCGCGCTAC
CGTTCAGGCGATACAATGGCTGACCCGAGCACCCGATCGCTTTAGCGCCGCCCAGTTC
AGCCCTTTACTCGAACACGAAAAATCATCAACGCAGATTATTAATCTGCTGGTATGG
AGCGGGCTGTGTGGCTGGATAAATCGCTTAAAAATCGCGTTGGGTGAGACATATTAA
CCCGGG

>CnoX_high

GAATTCGCGGCCGCTTCTAGATTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGTCCGTAGAAAATATTGTCAACATTAACGAAT
CTAACCTGCAACAGGTTCTTGAACAGTCGATGACCACTCCGGTGCTGTTCTATTTTTG
GTCTGAACGTAGCCAGCACTGTTTGCAGTTAACCCCAATTCTGGAAAGCCTCGCGGCG
CAGTACAACGGGCAATTTATTCTGGCGAAGCTGGACTGCGACGCGGAGCAGATGATT
GCCGCGCAGTTTGGTCTGCGTGCGATTCCGACCGTGTATCTGTTCCAGAACGGGCAAC
CGGTAGATGGCTTCCAGGGGCCGCAACCGGAAGAGGCGATCCGCGCCCTGCTGGATA
AAGTGCTGCCGCGCGAAGAAGAGCTGAAAGCGCAGCAGGCGATGCAACTGATGCAGG
AAAGCAATTACACCGATGCCCTGCCATTGCTGAAAGACGCCTGGCAGTTGTCGAATC
AGAACGGGGAGATCGGCCTGCTGCTGGCAGAAACGCTGATTGCGCTGAACCGTTCTG
AAGATGCGGAAGCGGTGCTGAAAACCATTCCGTTGCAGGATCAGGACACCCGCTACC
AGGGGCTGGTGGCGCAAATCGAACTGCTGAAGCAGGCGGCTGATACGCCGGAAATTC
AACAGTTGCAACAGCAGGTGGCGGAGAATCCAGAAGATGCCGCACTGGCGACGCAAC
TGGCGCTGCAACTGCATCAGGTTGGGCGCAATGAAGAGGCGCTGGAGTTGCTGTTCG
GGCATCTGCGTAAAGATCTCACCGCCGCAGACGGTCAGACGCGTAAAACGTTCCAGG
AGATCCTCGCTGCGCTGGGTACGGGTGATGCACTGGCGTCGAAGTATCGCCGCCAGCT
GTATGCATTGTTGTATTGACTGCAG

>CnoX_medium

GAATTCGCGGCCGCTTCTAGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGTCCGTAGAAAATATTGTCAACATTAACGAAT
CTAACCTGCAACAGGTTCTTGAACAGTCGATGACCACTCCGGTGCTGTTCTATTTTTG
GTCTGAACGTAGCCAGCACTGTTTGCAGTTAACCCCAATTCTGGAAAGCCTCGCGGCG
CAGTACAACGGGCAATTTATTCTGGCGAAGCTGGACTGCGACGCGGAGCAGATGATT
GCCGCGCAGTTTGGTCTGCGTGCGATTCCGACCGTGTATCTGTTCCAGAACGGGCAAC
CGGTAGATGGCTTCCAGGGGCCGCAACCGGAAGAGGCGATCCGCGCCCTGCTGGATA
AAGTGCTGCCGCGCGAAGAAGAGCTGAAAGCGCAGCAGGCGATGCAACTGATGCAGG
AAAGCAATTACACCGATGCCCTGCCATTGCTGAAAGACGCCTGGCAGTTGTCGAATC
AGAACGGGGAGATCGGCCTGCTGCTGGCAGAAACGCTGATTGCGCTGAACCGTTCTG
AAGATGCGGAAGCGGTGCTGAAAACCATTCCGTTGCAGGATCAGGACACCCGCTACC
AGGGGCTGGTGGCGCAAATCGAACTGCTGAAGCAGGCGGCTGATACGCCGGAAATTC
AACAGTTGCAACAGCAGGTGGCGGAGAATCCAGAAGATGCCGCACTGGCGACGCAAC
TGGCGCTGCAACTGCATCAGGTTGGGCGCAATGAAGAGGCGCTGGAGTTGCTGTTCG
GGCATCTGCGTAAAGATCTCACCGCCGCAGACGGTCAGACGCGTAAAACGTTCCAGG
AGATCCTCGCTGCGCTGGGTACGGGTGATGCACTGGCGTCGAAGTATCGCCGCCAGCT
GTATGCATTGTTGTATTGACTGCAG

>SecB_high

GAATTCGCGGCCGCTTCTAGATTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGTCAGAACAAAACAACACTGAAATGACTTTCC
AGATCCAACGTATTTATACCAAGGATATCTCTTTCGAAGCGCCGAACGCGCCGCACGT
TTTCCAGAAAGATTGGCAACCAGAAGTTAAACTTGATCTGGATACGGCATCTTCCCA
ACTGGCAGATGACGTATACGAAGTGGTACTGCGTGTTACCGTAACGGCCTCTTTGGG
CGAAGAAACCGCGTTCCTGTGTGAAGTTCAGCAGGGCGGTATTTTCTCCATCGCGGG
TATCGAAGGCACCCAGATGGCGCATTGCCTGGGAGCATACTGCCCGAACATTCTGTTC
CCGTATGCTCGTGAGTGCATCACCAGCATGGTATCCCGCGGTACATTCCCGCAACTGA
ACCTTGCGCCGGTTAACTTCGATGCGCTGTTCATGAACTATTTGCAGCAGCAGGCTGG
CGAAGGTACTGAAGAACATCAGGATGCCTGACTGCAG

>SecB_medium

GAATTCGCGGCCGCTTCTAGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGTCAGAACAAAACAACACTGAAATGACTTTCC
AGATCCAACGTATTTATACCAAGGATATCTCTTTCGAAGCGCCGAACGCGCCGCACGT
TTTCCAGAAAGATTGGCAACCAGAAGTTAAACTTGATCTGGATACGGCATCTTCCCA
ACTGGCAGATGACGTATACGAAGTGGTACTGCGTGTTACCGTAACGGCCTCTTTGGG
CGAAGAAACCGCGTTCCTGTGTGAAGTTCAGCAGGGCGGTATTTTCTCCATCGCGGG
TATCGAAGGCACCCAGATGGCGCATTGCCTGGGAGCATACTGCCCGAACATTCTGTTC
CCGTATGCTCGTGAGTGCATCACCAGCATGGTATCCCGCGGTACATTCCCGCAACTGA
ACCTTGCGCCGGTTAACTTCGATGCGCTGTTCATGAACTATTTGCAGCAGCAGGCTGG
CGAAGGTACTGAAGAACATCAGGATGCCTGACTGCAG

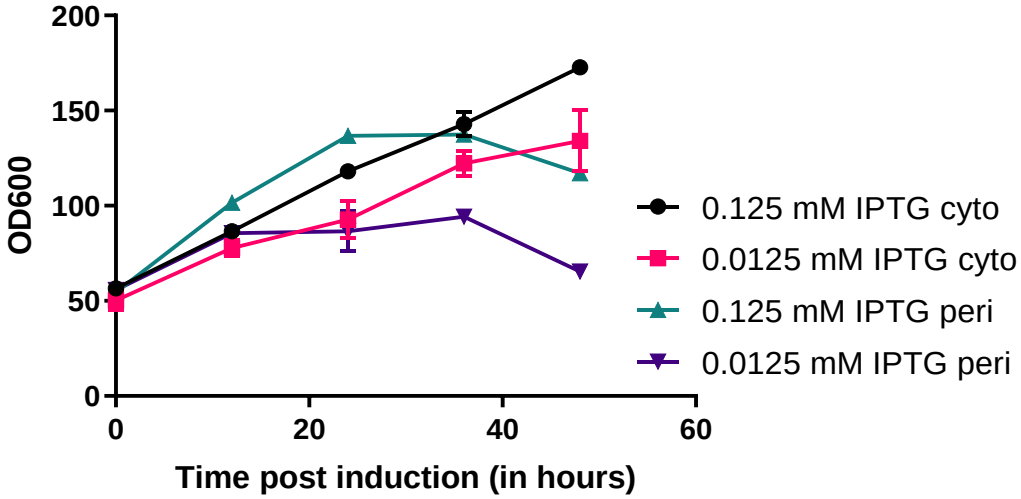
>Spy_medium

GAATTCGCGGCCGCTTCTAGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCTC
TAGAGAAAGAGGGGACAAACTAGATGCGTAAATTAACTGCACTGTTTGTTGCCTCTA
CCCTGGCTCTTGGCGCGGCTAACCTGGCCCATGCCGCAGACACCACTACCGCAGCACC
GGCTGACGCGAAGCCGATGATGCACCACAAAGGCAAGTTCGGTCCGCATCAGGACAT
GATGTTCAAAGACCTGAACCTGACCGACGCGCAGAAACAGCAGATCCGCGAAATCAT
GAAAGGCCAGCGTGACCAGATGAAACGTCCGCCGCTGGAAGAACGCCGCGCAATGCA
TGACATCATTGCCAGCGATACCTTCGATAAAGTAAAAGCTGAAGCGCAGATCGCAAA
AATGGAAGAACAGCGCAAAGCTAACATGCTGGCGCACATGGAAACCCAGAACAAAAT
TTACAACATCCTGACGCCGGAACAGAAAAAGCAATTTAATGCTAATTTTGAGAAGCG
TCTGACAGAACGTCCAGCGGCAAAAGGTAAAATGCCTGCAACTGCTGAATAACTGCA
G

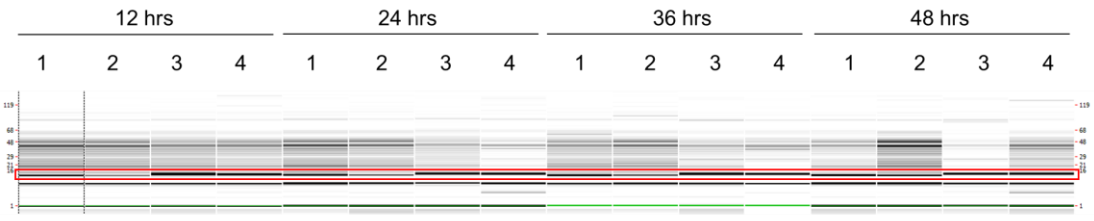
>Spy_low

GAATTCGCGGCCGCTTCTAGACTGATAGCTAGCTCAGTCCTAGGGATTATGCTAGCCT
AGAGAAAGAGGGGACAAACTAGATGCGTAAATTAACTGCACTGTTTGTTGCCTCTAC
CCTGGCTCTTGGCGCGGCTAACCTGGCCCATGCCGCAGACACCACTACCGCAGCACCG
GCTGACGCGAAGCCGATGATGCACCACAAAGGCAAGTTCGGTCCGCATCAGGACATG
ATGTTCAAAGACCTGAACCTGACCGACGCGCAGAAACAGCAGATCCGCGAAATCATG
AAAGGCCAGCGTGACCAGATGAAACGTCCGCCGCTGGAAGAACGCCGCGCAATGCAT
GACATCATTGCCAGCGATACCTTCGATAAAGTAAAAGCTGAAGCGCAGATCGCAAAA
ATGGAAGAACAGCGCAAAGCTAACATGCTGGCGCACATGGAAACCCAGAACAAAATT
TACAACATCCTGACGCCGGAACAGAAAAAGCAATTTAATGCTAATTTTGAGAAGCGT
CTGACAGAACGTCCAGCGGCAAAAGGTAAAATGCCTGCAACTGCTGAATAACTGCAG

Appendix 3. Growth curve of high cell density E. coli cultures expressing the scFv in the cytoplasm (cyto P11-scFv) or periplasm (peri P11-scFv+sp) induced with 0.125 mM IPTG or 0.0125 mM IPTG. Bacterial concentration was monitored using OD_{600}. Error bars = 1 standard deviation, N=3 for all samples except cells inducing expression of the scFv in the periplasm with 0.0125 mM IPTG where N=2.

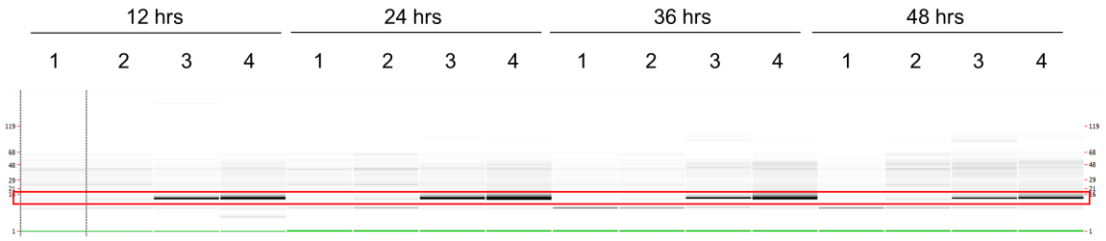


Appendix 4. Capillary electrophoresis of cells overexpressing the scFv in the cytoplasm (scFv) or periplasm (scFv+sp) induced with 0.125 mM IPTG or 0.0125 mM IPTG, 12-, 24-, 36- and 48-hours post-induction under Ambr fermentation conditions. Samples were normalised to the bacterial concentration (OD_{600}). The scFv has a molecular weight of 12 kDa-14 kDa and is circled in red. Capillary electrophoresis does not have high enough resolution to separate scFv bands with and without the OmpA signal peptide.



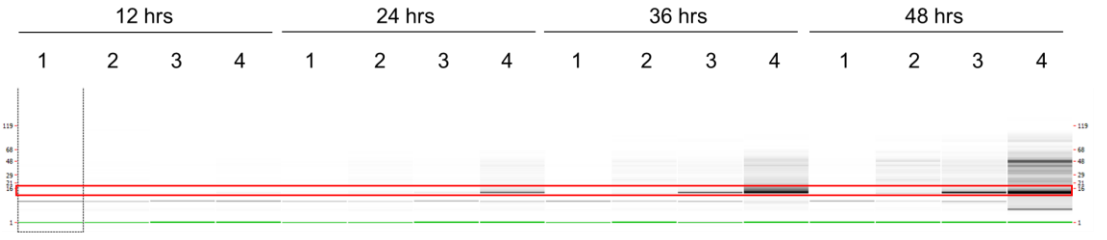
1	ScFv (cyto) + 0.125 mM IPTG
2	ScFv (cyto) + 0.0125 mM IPTG
3	ScFv+sp (peri) + 0.125 mM IPTG
4	ScFv+sp (peri) + 0.0125 mM IPTG

Appendix 5. Capillary electrophoresis of supernatant samples when overexpressing the scFv in the cytoplasm (scFv) or periplasm (scFv+sp) induced with 0.125 mM IPTG or 0.0125 mM IPTG, 12-, 24-, 36- and 48-hours post-induction under Ambr fermentation conditions. Samples were normalised to the bacterial concentration (OD_{600}). The scFv has a molecular weight of 12 kDa-14 kDa and is circled in red. Capillary electrophoresis does not have high enough resolution to separate scFv bands with and without the OmpA signal peptide.



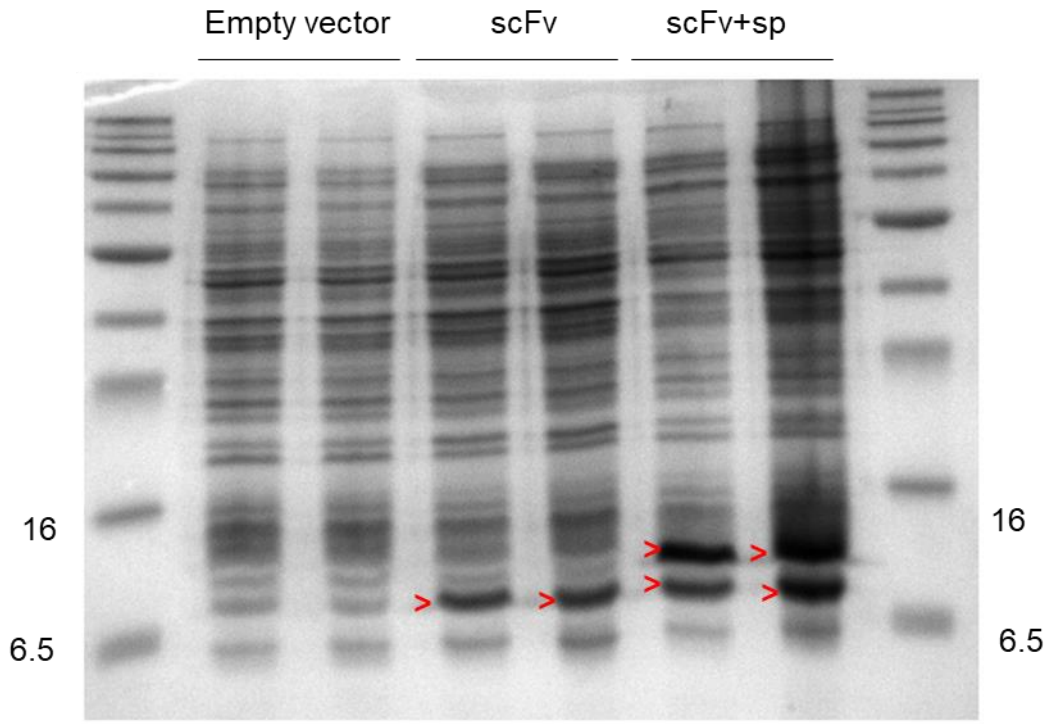
1	ScFv (cyto) + 0.125 mM IPTG
2	ScFv (cyto) + 0.0125 mM IPTG
3	ScFv+sp (peri) + 0.125 mM IPTG
4	ScFv+sp (peri) + 0.0125 mM IPTG

Appendix 6. Capillary electrophoresis of supernatant fractions when overexpressing the scFv in the cytoplasm (scFv) or periplasm (scFv+sp) induced with 0.125 mM IPTG or 0.0125 mM IPTG, 12-, 24-, 36- and 48-hours post-induction under Ambr fermentation conditions. Samples were normalised to the bacterial concentration (OD_{600}). The scFv has a molecular weight of 12 kDa-14 kDa and is circled in red. Capillary electrophoresis does not have high enough resolution to separate scFv bands with and without the OmpA signal peptide.

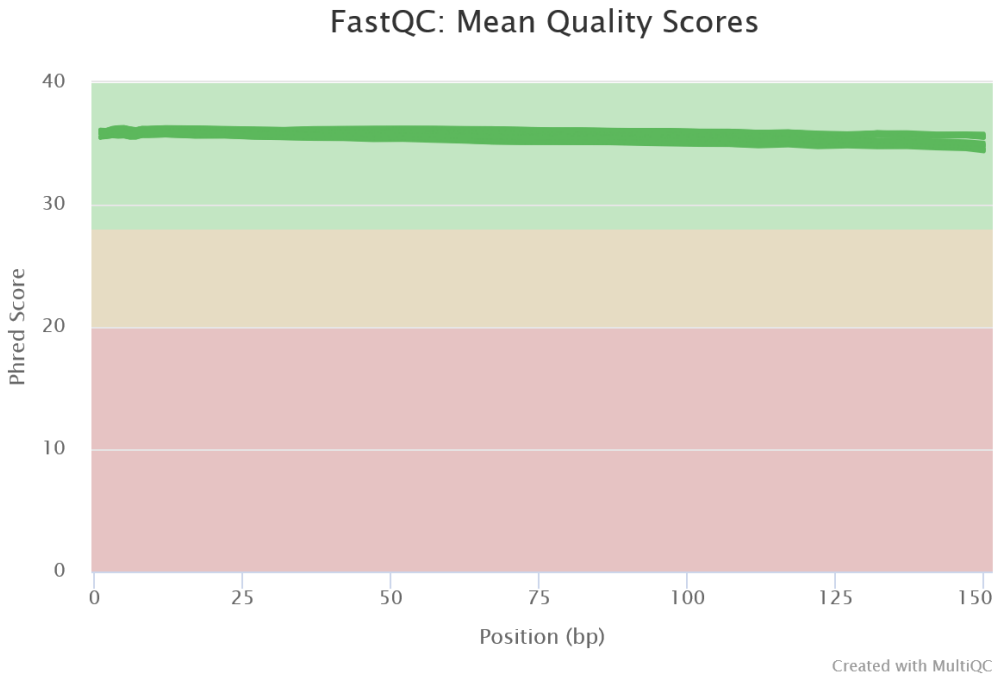


1	ScFv (cyto) + 0.125 mM IPTG
2	ScFv (cyto) + 0.0125 mM IPTG
3	ScFv+sp (peri) + 0.125 mM IPTG
4	ScFv+sp (peri) + 0.0125 mM IPTG

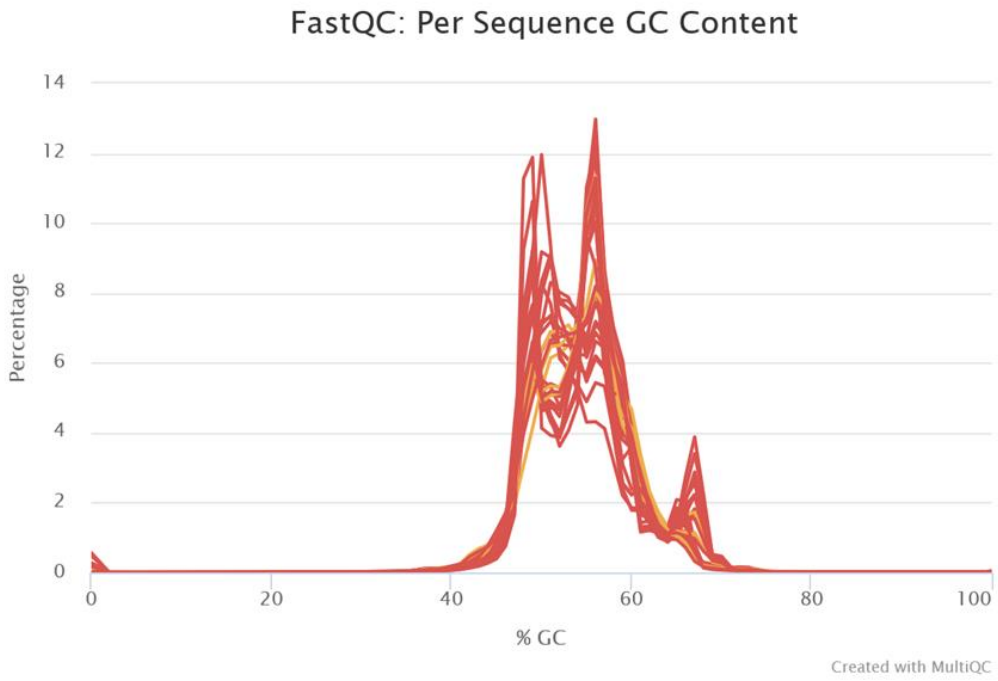
Appendix 7. SDS-PAGE of whole cell samples used for TEM analysis (biological replicate 2 and 3). Samples containing an empty vector plasmid, expressing the scFv in the cytoplasm (scFv) and periplasm (scFv+sp) were grown in shake flasks and induced with 0.125 mM IPTG mid log phase. Samples are normalised based on OD_{600}. The scFv has a molecular weight of



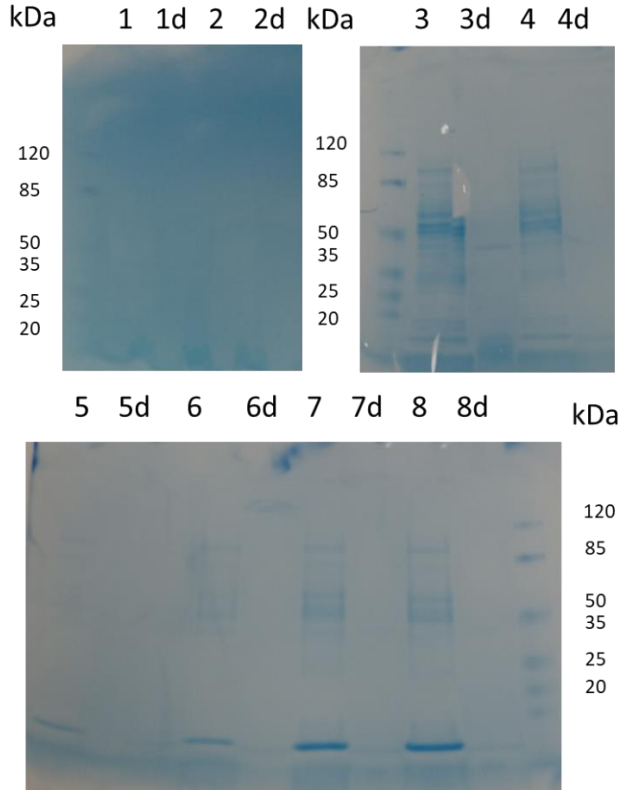
Appendix 8. Per base sequence quality plot of RNA-seq sequencing data. Average sequencing quality for each sample (forward and reverse) is indicated by the green lines. Green represents good sequence quality, yellow intermediate, and red bad sequence quality. The plot was generated using MultiQC [241]***.***



Appendix 9. Plot of QC content per read from RNA-seq samples. The QC content per read for each sample (forward and reverse) is represented by each line. Data should be normally distributed around 50% GC content. Green lines represent good, yellow medium and red a bad GC content per read. This plot was generated using MultiQC [241].

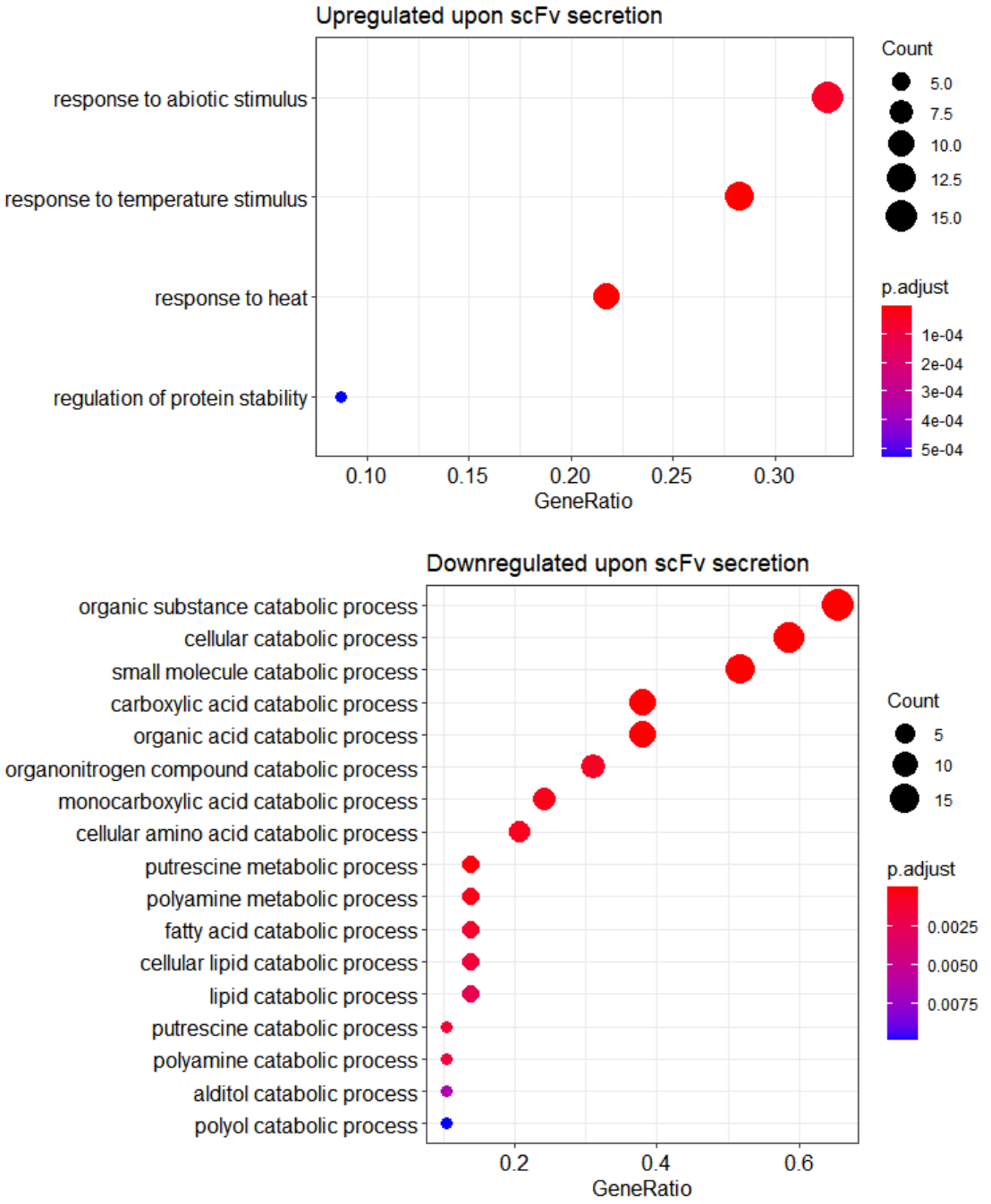


Appendix 10. SDS-PAGE of proteomics samples overexpressing an scFv in the cytoplasm (cyto - scFv) and periplasm (peri - scFv+sp) induced with different concentrations of inducer (high (0.125 mM IPTG) and low (0.0125 mM IPTG)) pre and post trypsin digestion overnight (d). Numbers (1, 2 and 3) indicate biological replicates. Cells were lysed using bead beating. Protein concentrations were analysed using a Bradford assay and 10 µg were loaded onto an SDS page gel. No bans are present in samples 1 and 2 which could be due to multiple factors including issues with the gel, strain and samples preparation as well as digestion of samples. Bands were still present in digested sample 3 and therefore an extra trypsin digestion step was conducted.

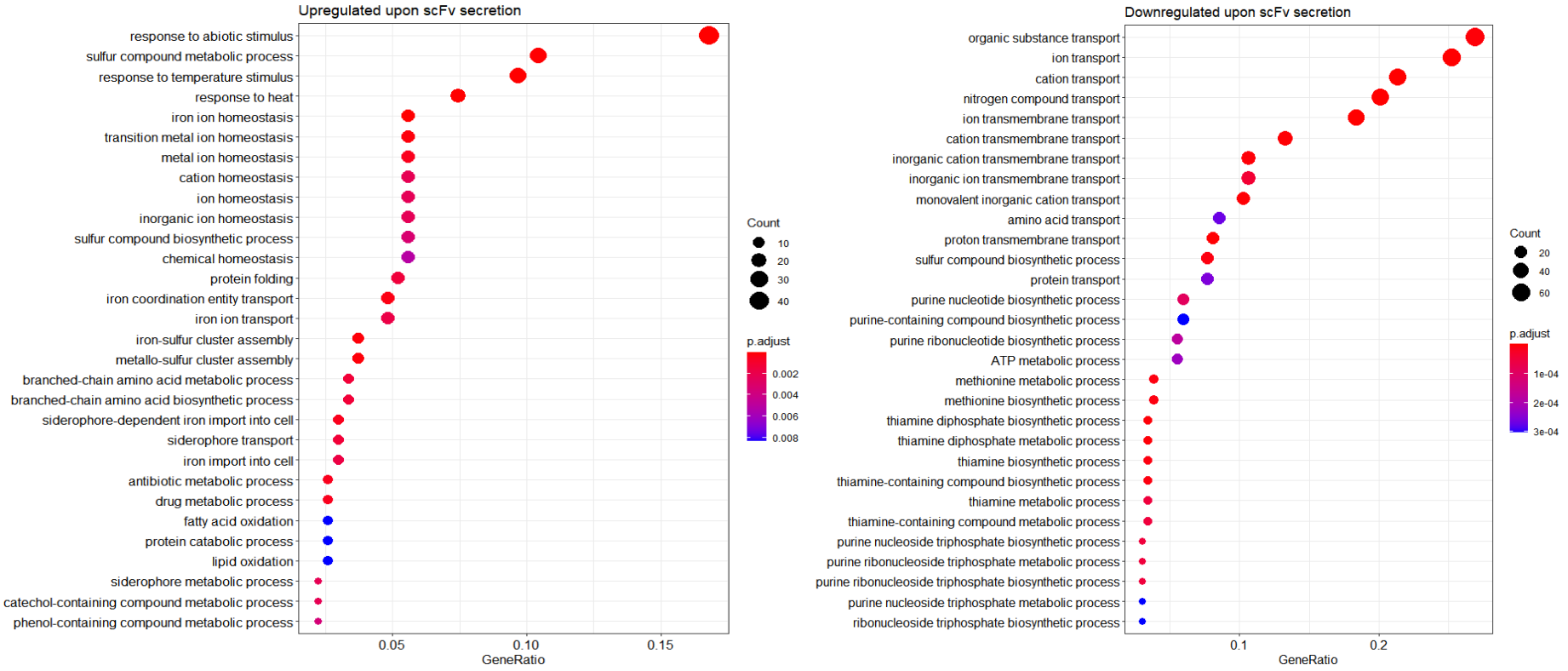


1	scFv (cyto) high 0.125 mM IPTG 1
2	scFv (cyto) high 0.125 mM IPTG 3
3	scFv (cyto) low 0.0125 mM IPTG 1
4	scFv (cyto) low 0.0125 mM IPTG 2
5	scFv+sp (peri) high 0.125 mM IPTG 1
6	scFv+sp (peri) high 0.125 mM IPTG 2
7	scFv+sp (peri) low 0.0125 mM IPTG 1
8	scFv+sp (peri) low 0.0125 mM IPTG 2

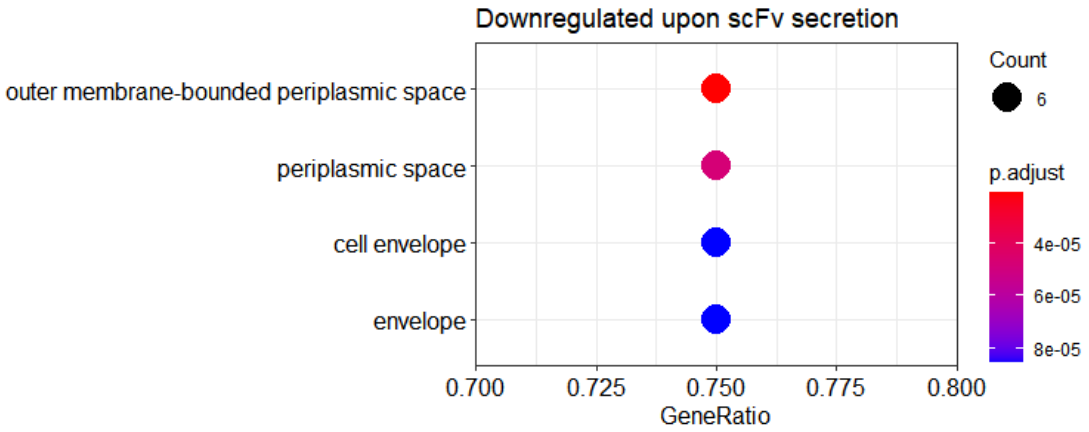
Appendix 11. Dot plot representing the 30 most enriched gene ontology terms when using proteomics to compare cells overexpressing the scFv in the periplasm and cytoplasm while induced with 0.125 mM IPTG. Enrichment was conducted using molecular function MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package ClusterProfiler [255].



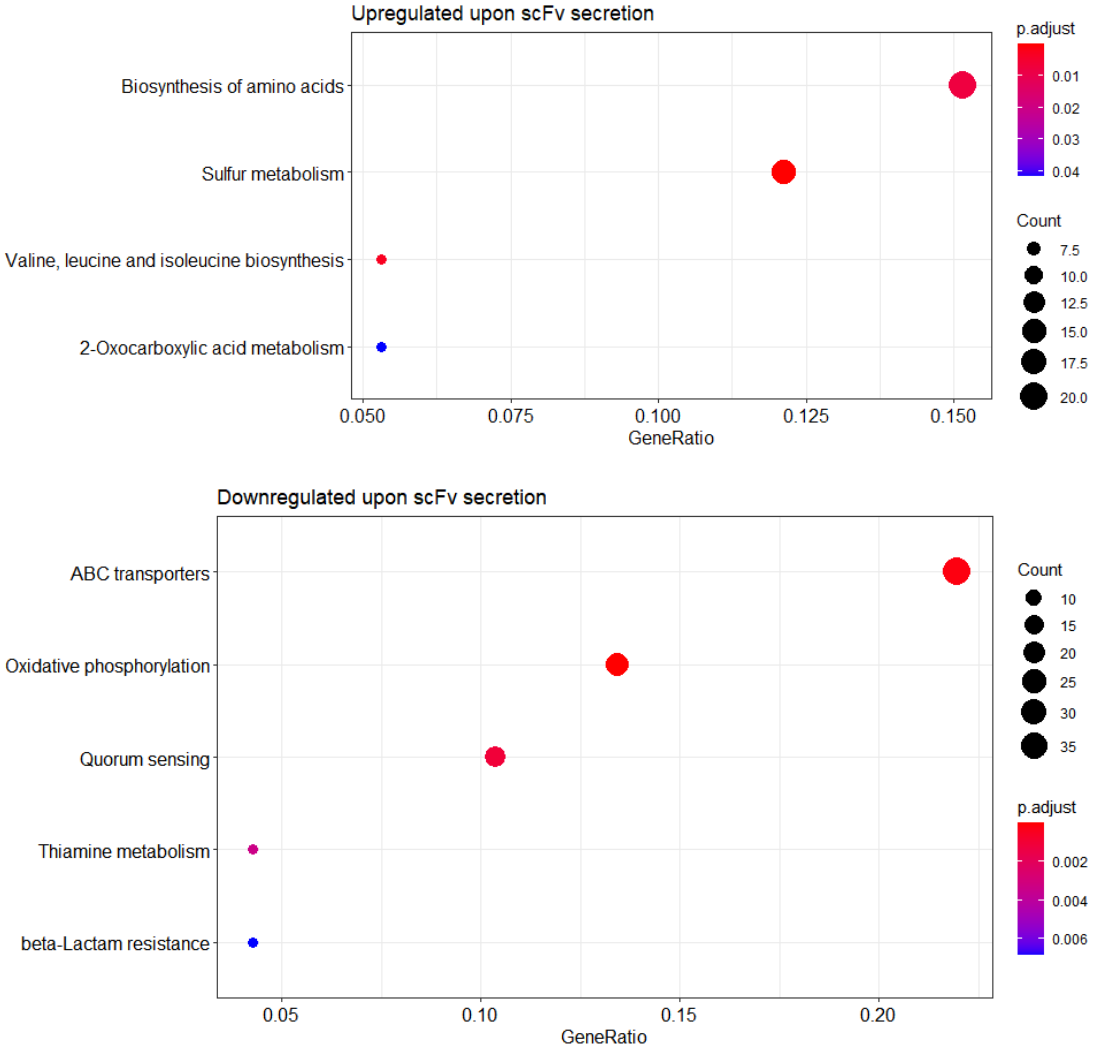
Appendix 12. Dot plot representing the 30 most enriched gene ontology terms when using transcriptomics to compare cells overexpressing the scFv in the periplasm and cytoplasm while induced with 0.0125 mM IPTG. Enrichment was conducted using molecular function MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package cluterProfiler [255]***.***



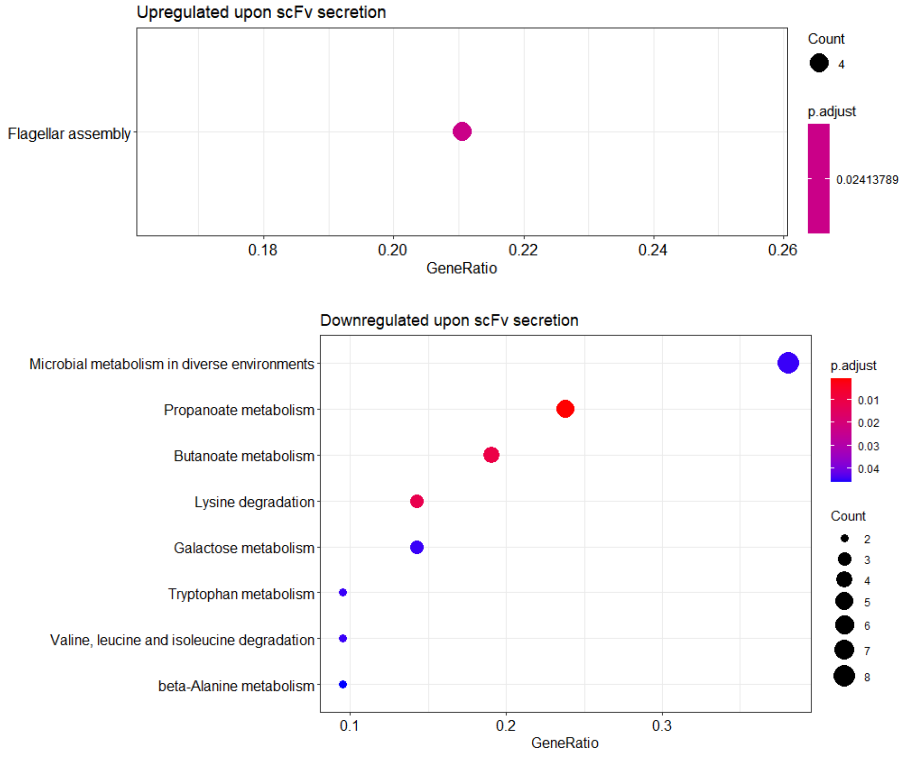
Appendix 13. Dot plot representing the 30 most enriched gene ontology terms when using proteomics to compare cells overexpressing the scFv in the periplasm and cytoplasm while induced with 0.0125 mM IPTG. Enrichment was conducted using molecular function MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255]***. No GO terms were enriched amongst proteins that were upregulated upon scFv expression.***



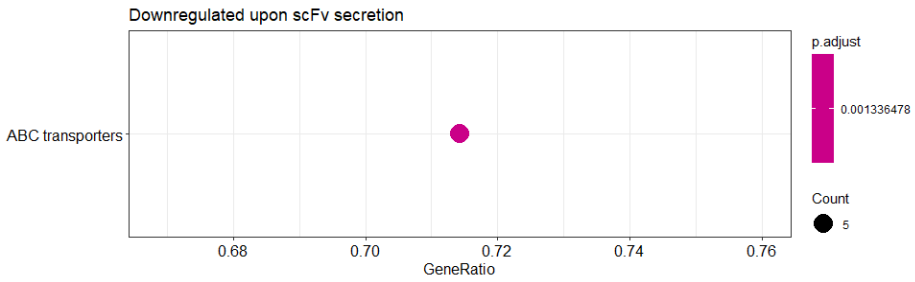
Appendix 14. Dot plot representing enriched KEGG terms when using transcriptomics to compare cells overexpressing the scFv in the periplasm and cytoplasm while induced with 0.0125 mM IPTG. The size of each dot represents the number of differentially expressed genes in said KEGG pathway and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes within the KEGG pathway. Pathways are considered enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package cluterProfiler [255]***.***



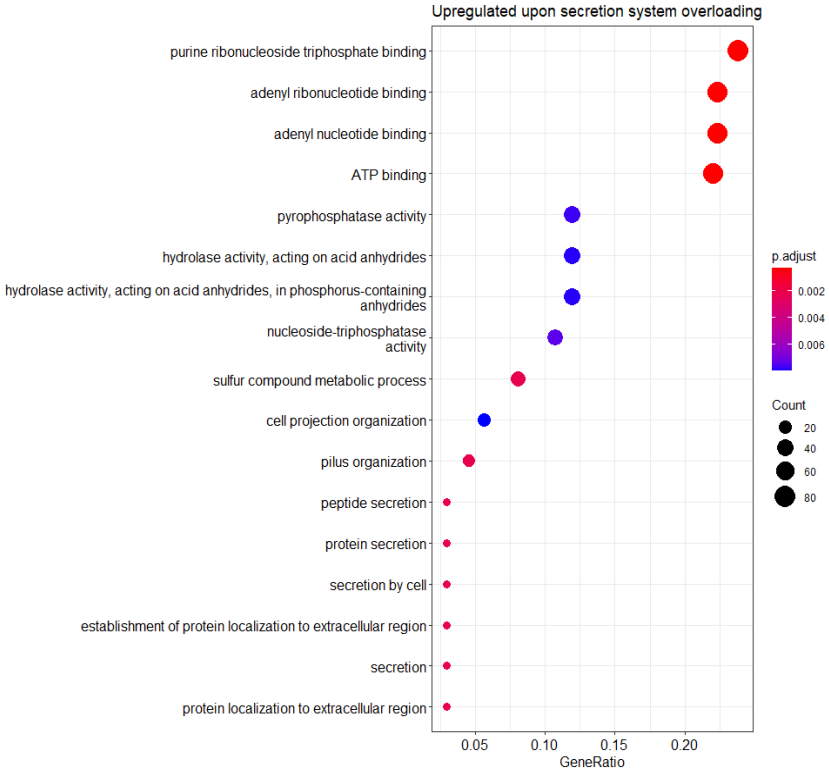
Appendix 15. Dot plot representing enriched KEGG terms when using proteomics to compare cells overexpressing the scFv in the periplasm and cytoplasm while induced with 0.125 mM IPTG. The size of each dot represents the number of differentially expressed genes in said KEGG pathway and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes within the KEGG pathway. Pathways are considered enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package cluterProfiler [255]***.***

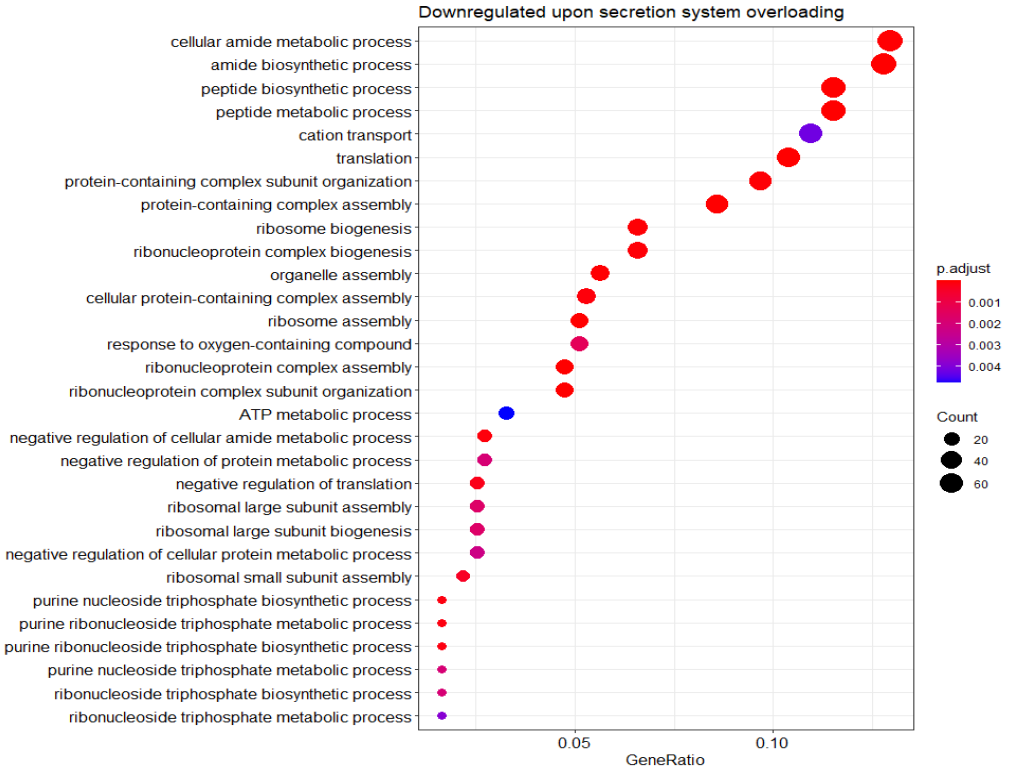


Appendix 16. Dot plot representing enriched KEGG terms when using proteomics to compare cells overexpressing the scFv in the periplasm and cytoplasm while induced with 0.0125 mM IPTG.The size of each dot represents the number of differentially expressed genes in said KEGG pathway and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes within the KEGG pathway. Pathways are considered enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package cluterProfiler [255]***.***

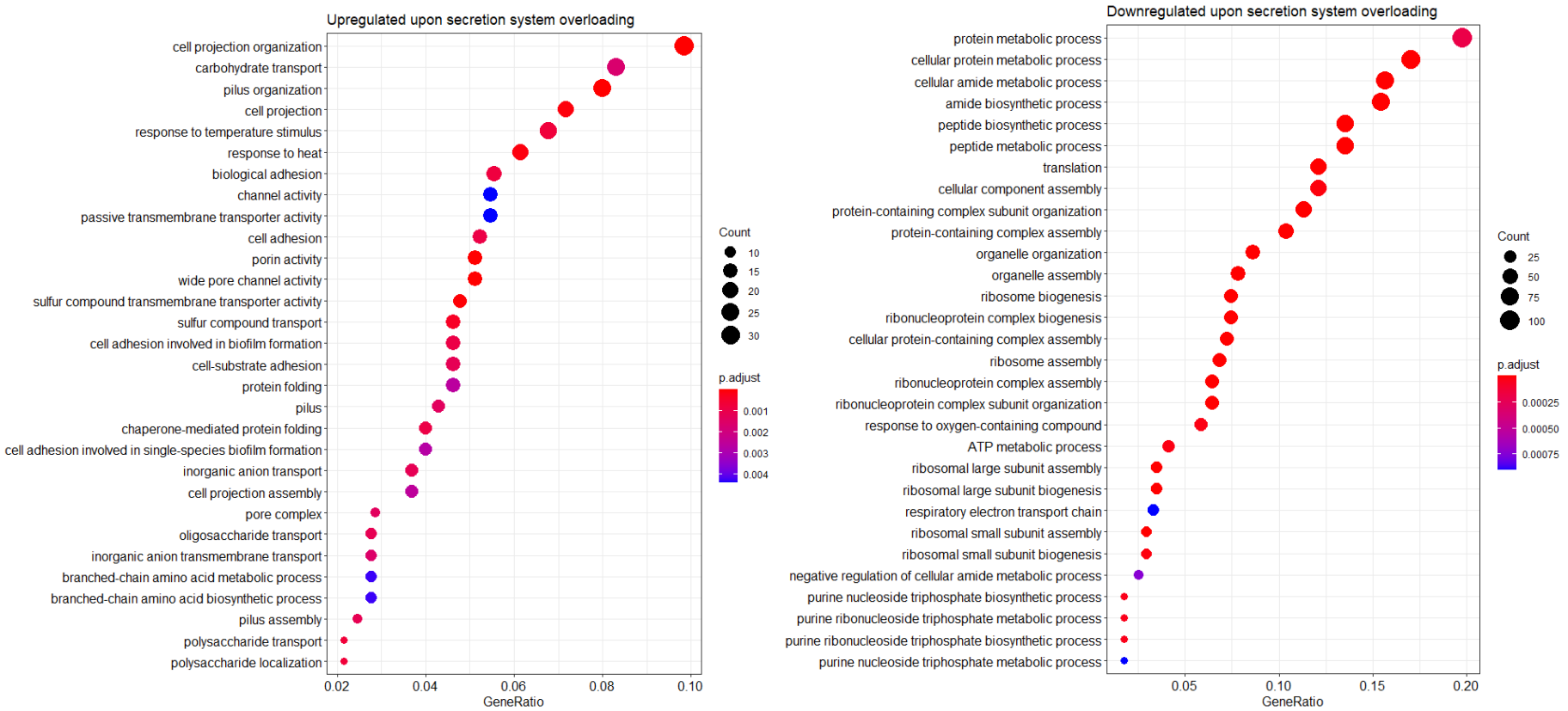


Appendix 17. Dot plot representing the 30 most enriched gene ontology terms when using transcriptomics to compare cells expressing the scFv in the periplasm when induced with high (0.125 mM secretion system overloading) and low (0.0125 mM) IPTG, 12 hours post induction. Enrichment was conducted using molecular function (MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255].

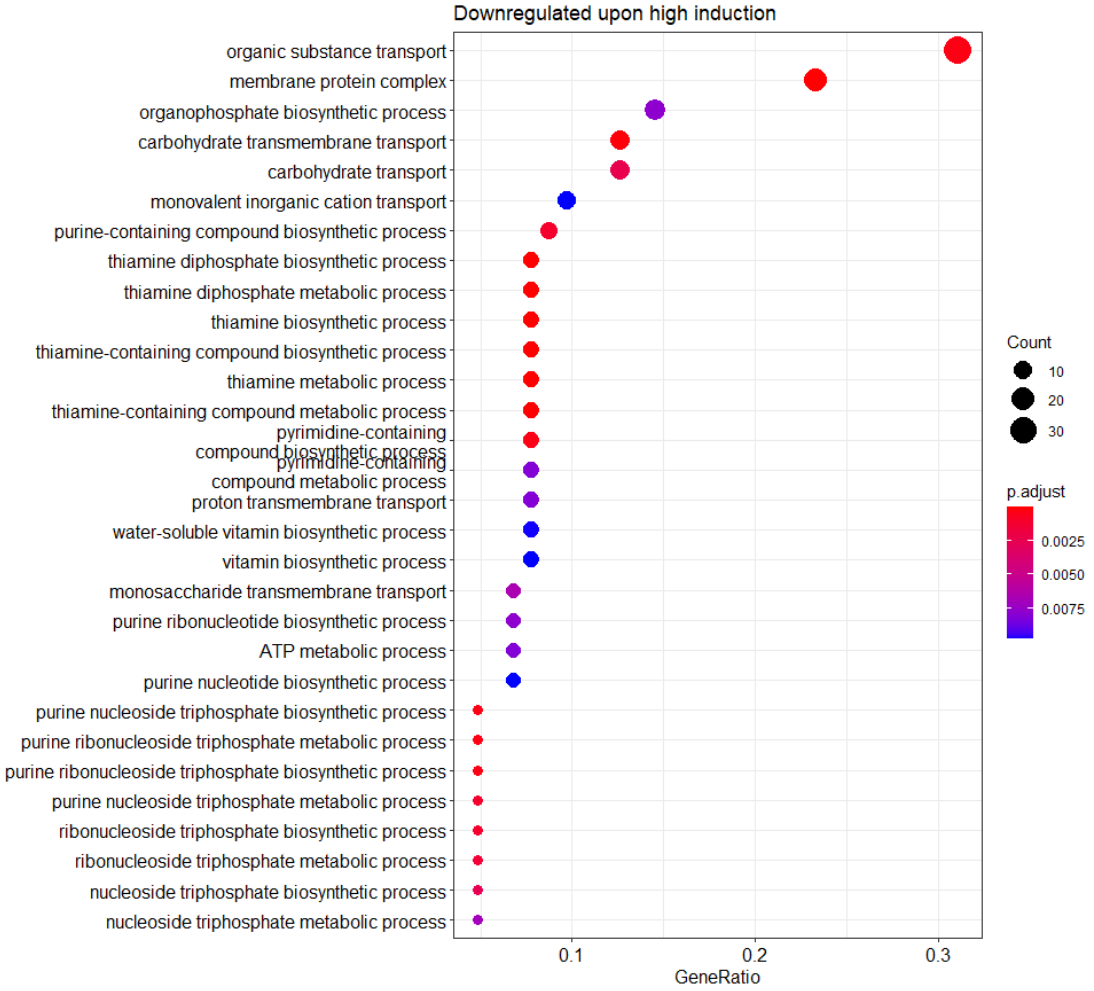




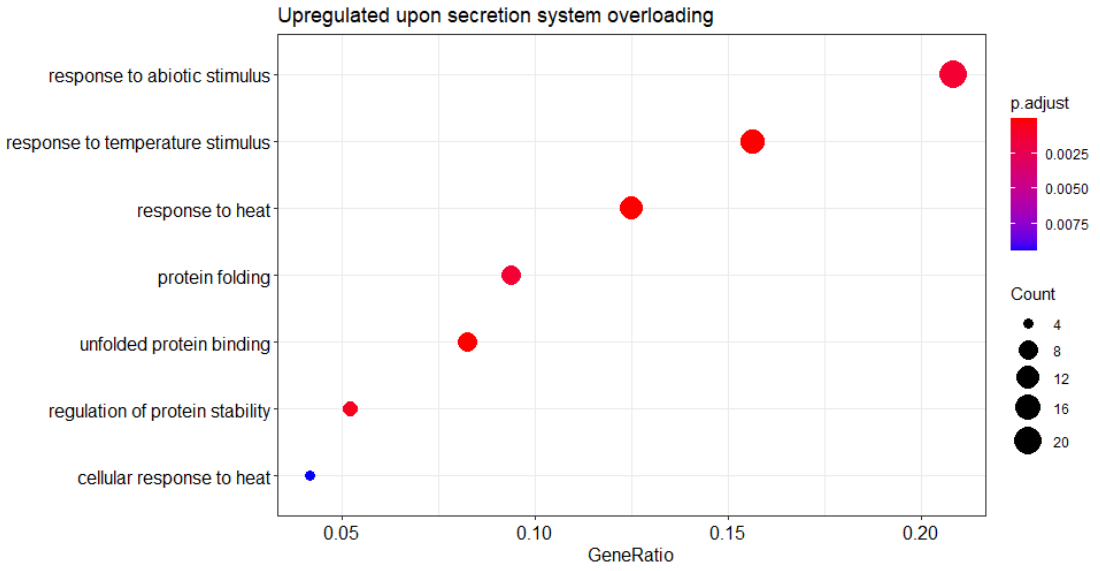
Appendix 18. Dot plot representing the 30 most enriched gene ontology terms when using transcriptomics to compare cells expressing the scFv in the periplasm when induced with high (0.125 mM secretion system overloading) and low (0.0125 mM) IPTG, 24 hours post induction. Enrichment was conducted using molecular function (MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255].



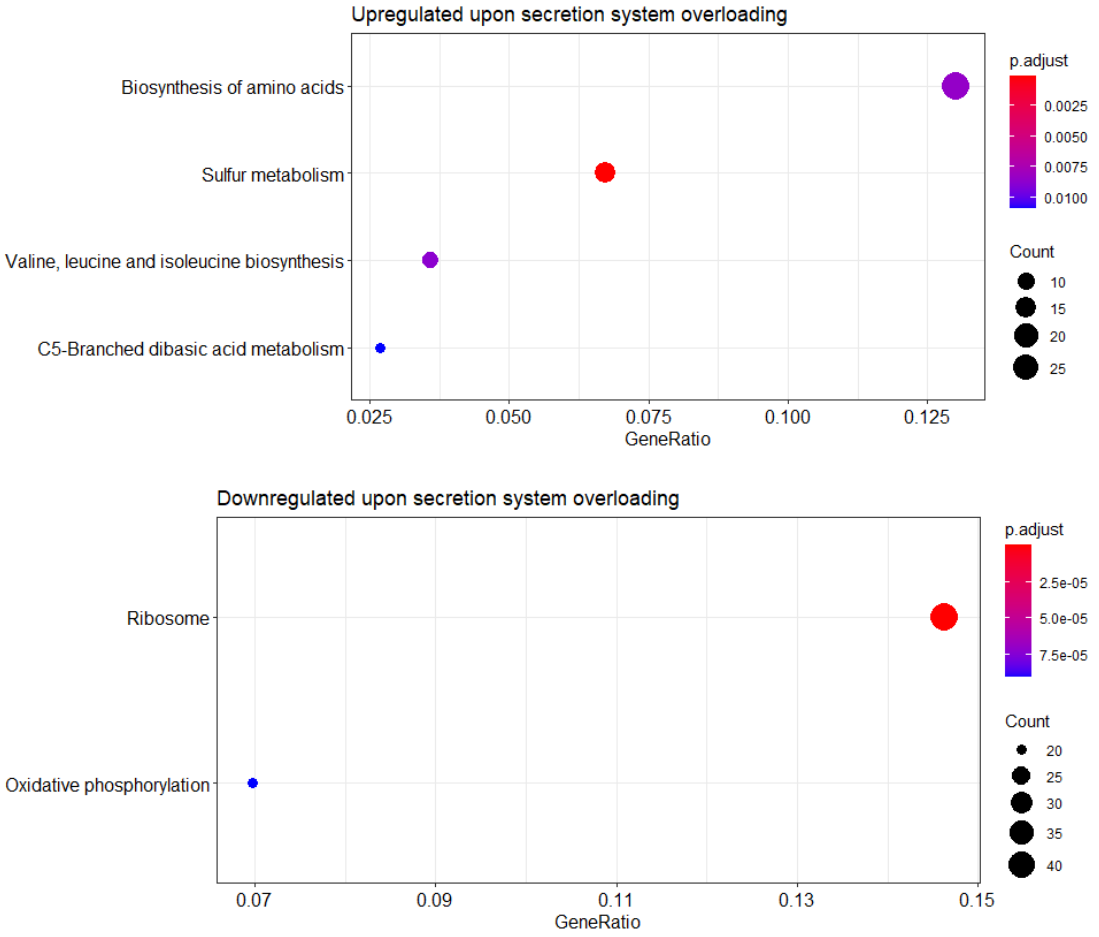
Appendix 19. Dot plot representing the 30 most enriched gene ontology terms when using proteomics to compare cells expressing the scFv in the cytoplasm when induced with high (0.125 mM) and low (0.0125 mM) IPTG, 12 hours post induction. Enrichment was conducted using molecular function (MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255].



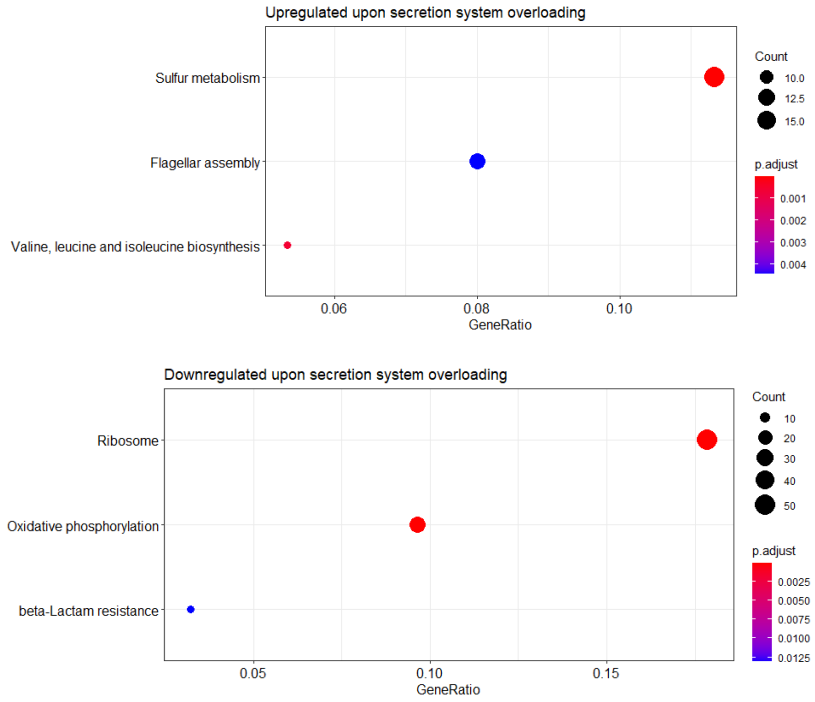
Appendix 20. Dot plot representing the 30 most enriched gene ontology terms when using proteomics to compare cells expressing the scFv in the periplasm when induced with high (0.125 mM secretion system overloading) and low (0.0125 mM) IPTG, 24 hours post induction. Enrichment was conducted using molecular function (MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255]***.***



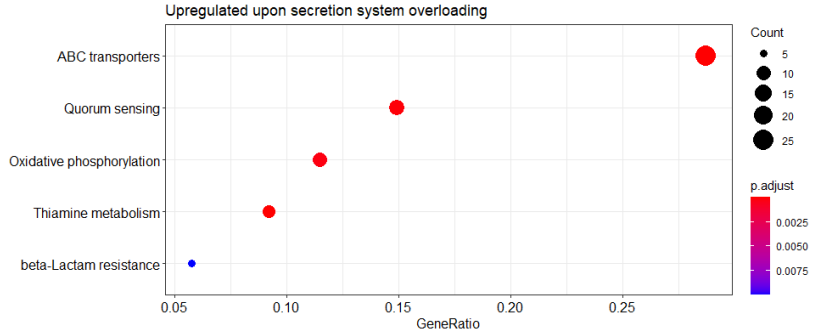
Appendix 21. Dot plot representing enriched KEGG terms when using transcriptomics to compare cells overexpression of the scFv in the periplasm when induced with high (0.125 mM IPTG secretion system overloading) and low (0.0125 mM IPTG) concentrations of inducer 12 hrs post induction. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255]***. The size of each dot represents the number of differentially expressed genes in said KEGG pathway and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes within the KEGG pathway. Pathways are considered enriched if they have an adjusted p-value < 0.05.***



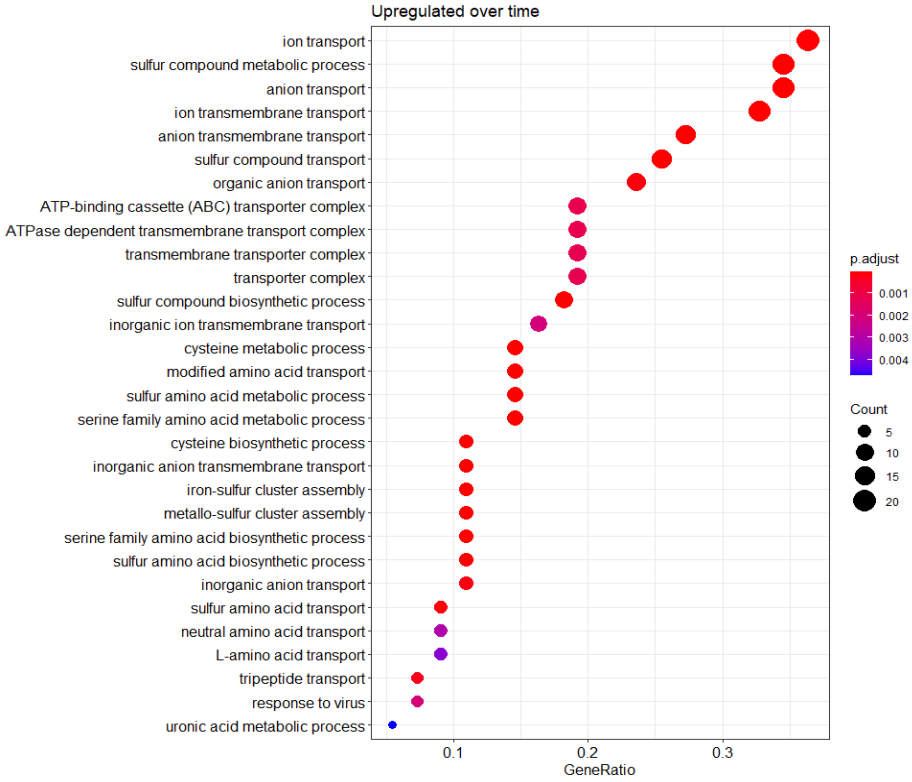
Appendix 22. Dot plot representing enriched KEGG terms when using transcriptomics to compare cells overexpression of the scFv in the periplasm when induced with high (0.125 mM IPTG secretion system overloading) and low (0.0125 mM IPTG) concentrations of inducer 24 hrs post induction. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255]***. The size of each dot represents the number of differentially expressed genes in said KEGG pathway and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes within the KEGG pathway. Pathways are considered enriched if they have an adjusted p-value < 0.05.***

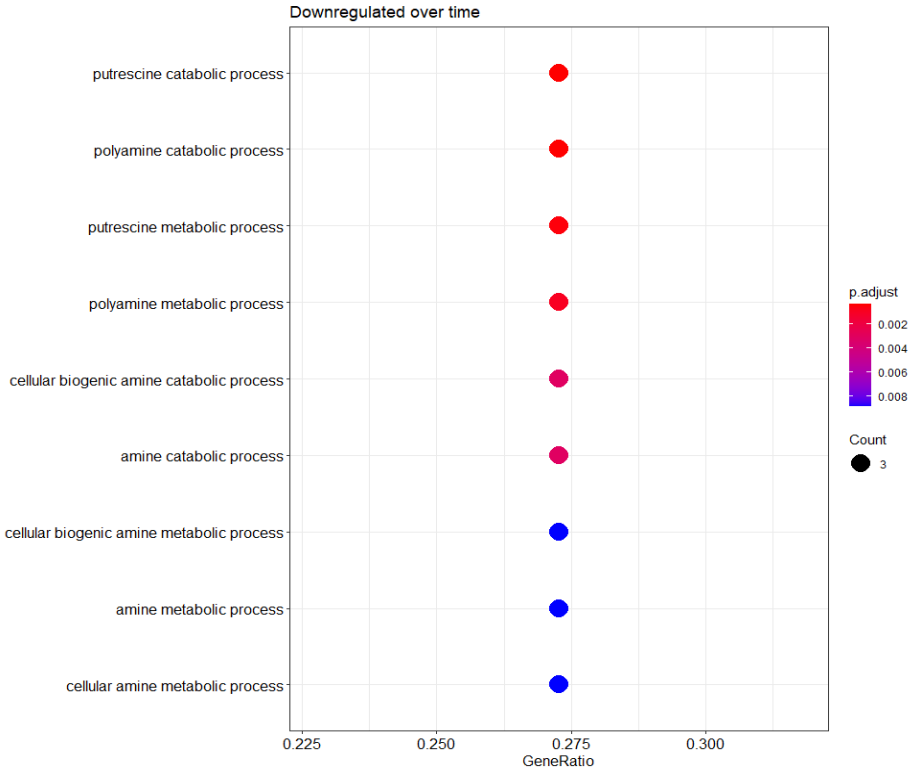


Appendix 23. Dot plot representing enriched KEGG terms when using proteomics to compare cells overexpression of the scFv in the cytoplasm when induced with high (0.125 mM IPTG secretion system overloading) and low (0.0125 mM IPTG) concentrations of inducer 24hrs post induction. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255]***. The size of each dot represents the number of differentially expressed genes in said KEGG pathway and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes within the KEGG pathway. Pathways are considered enriched if they have an adjusted p-value < 0.05.***

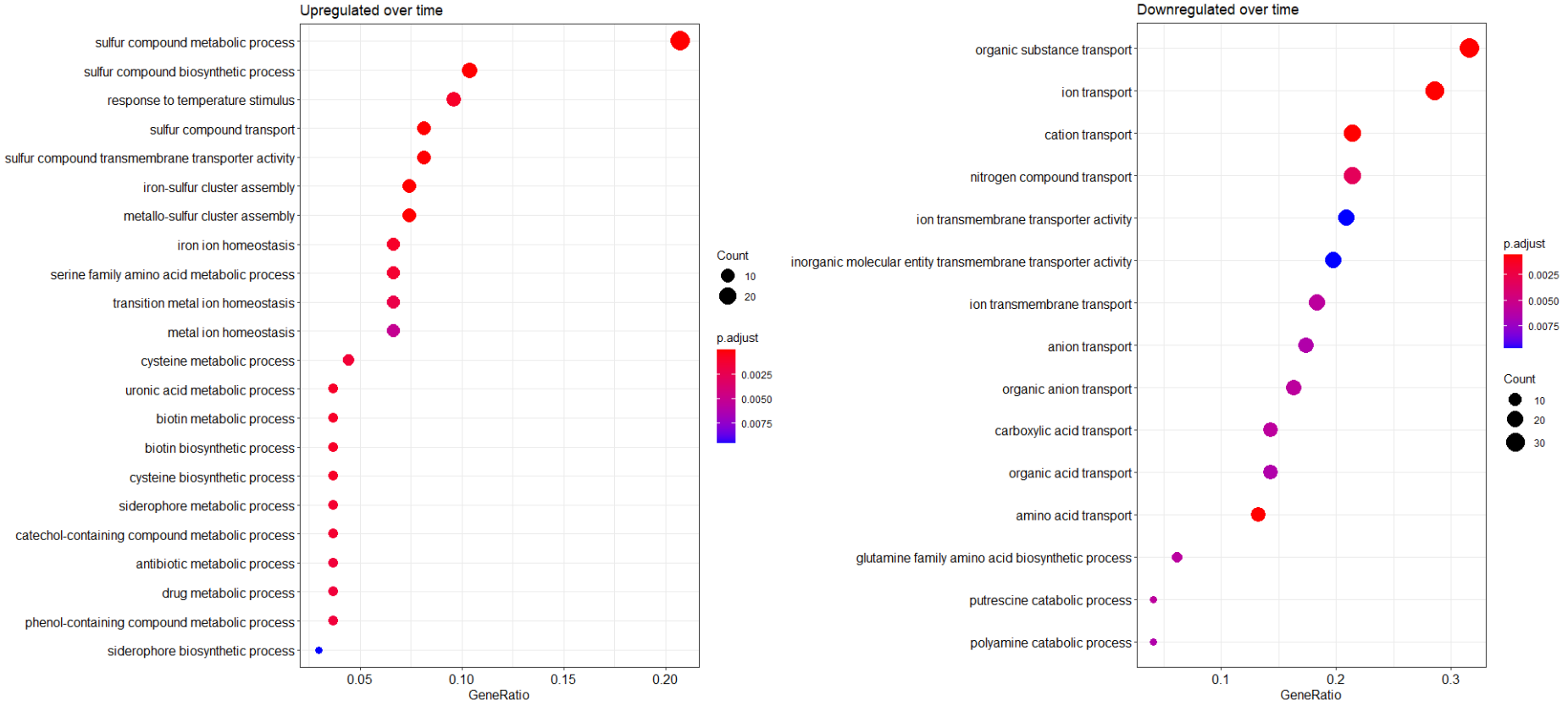


Appendix 24. Dot plot representing the 30 most enriched gene ontology terms when using transcriptomics to compare cells expressing the scFv in the periplasm (0.125 mM IPTG) 12 and 24 hours post induction. Enrichment was conducted using molecular function (MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255].

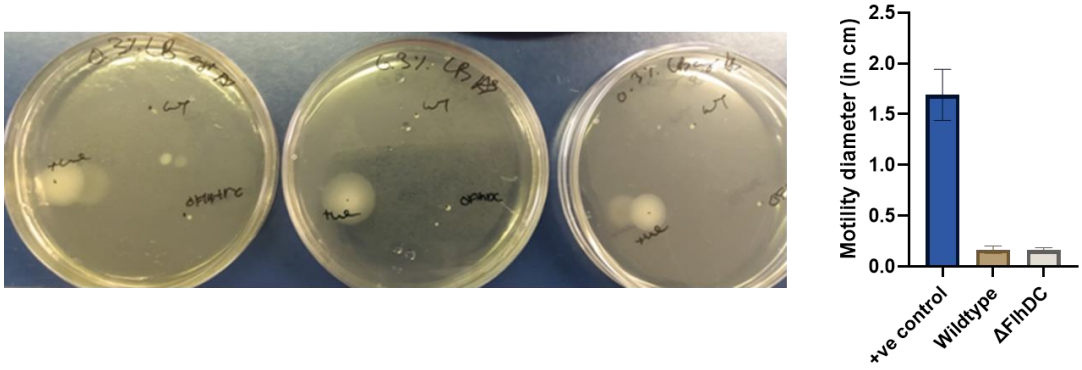




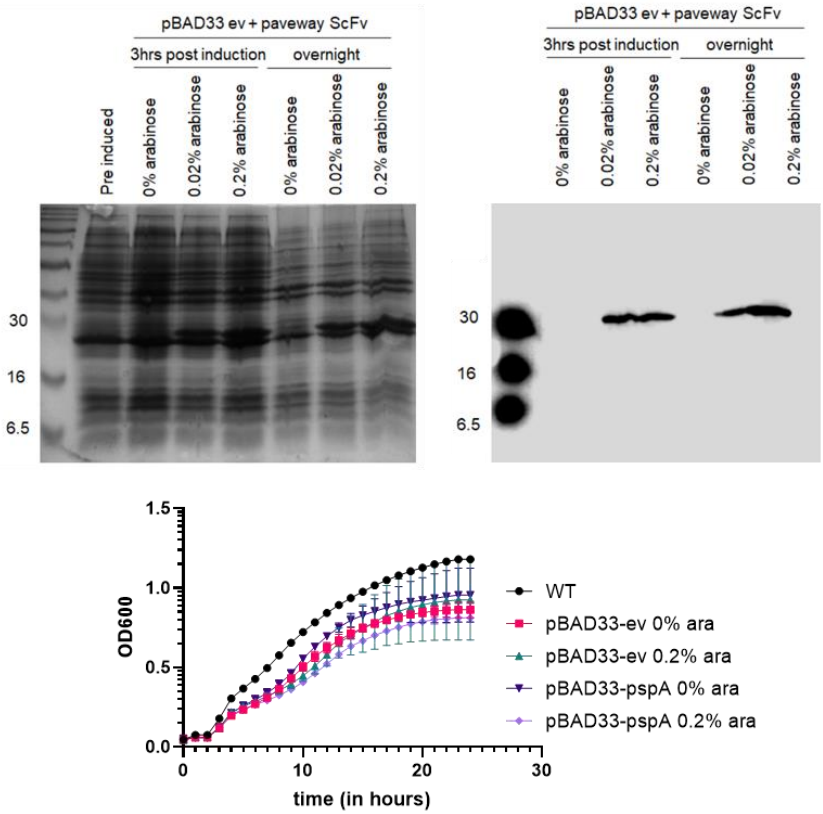
Appendix 25. Dot plot representing the 30 most enriched gene ontology terms when using transcriptomics to compare cells expressing the scFv in the periplasm (0.0125 mM IPTG) 12 and 24 hours post induction. Enrichment was conducted using molecular function (MF), cellular component (CC) and biological process (BP) gene ontology terms. The size of each dot represents the number of differentially expressed genes and the gene ratio represents the ratio of differentially expressed genes compared to the total number of genes with said GO term. GO terms are enriched if they have an adjusted p-value < 0.05. Enrichment and construction of dot plots was conducted using the R package clusterProfiler [255].



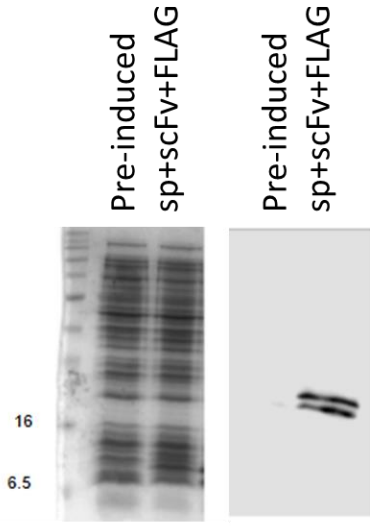
Appendix 26. Motility of wildtype and ΔflhDC strains compared to a positive control (E. coli MC1000). Motility was measured by spotting colonies onto motility agar and measuring diameter of bacterial swimming. Error bars represent 1 standard deviation.



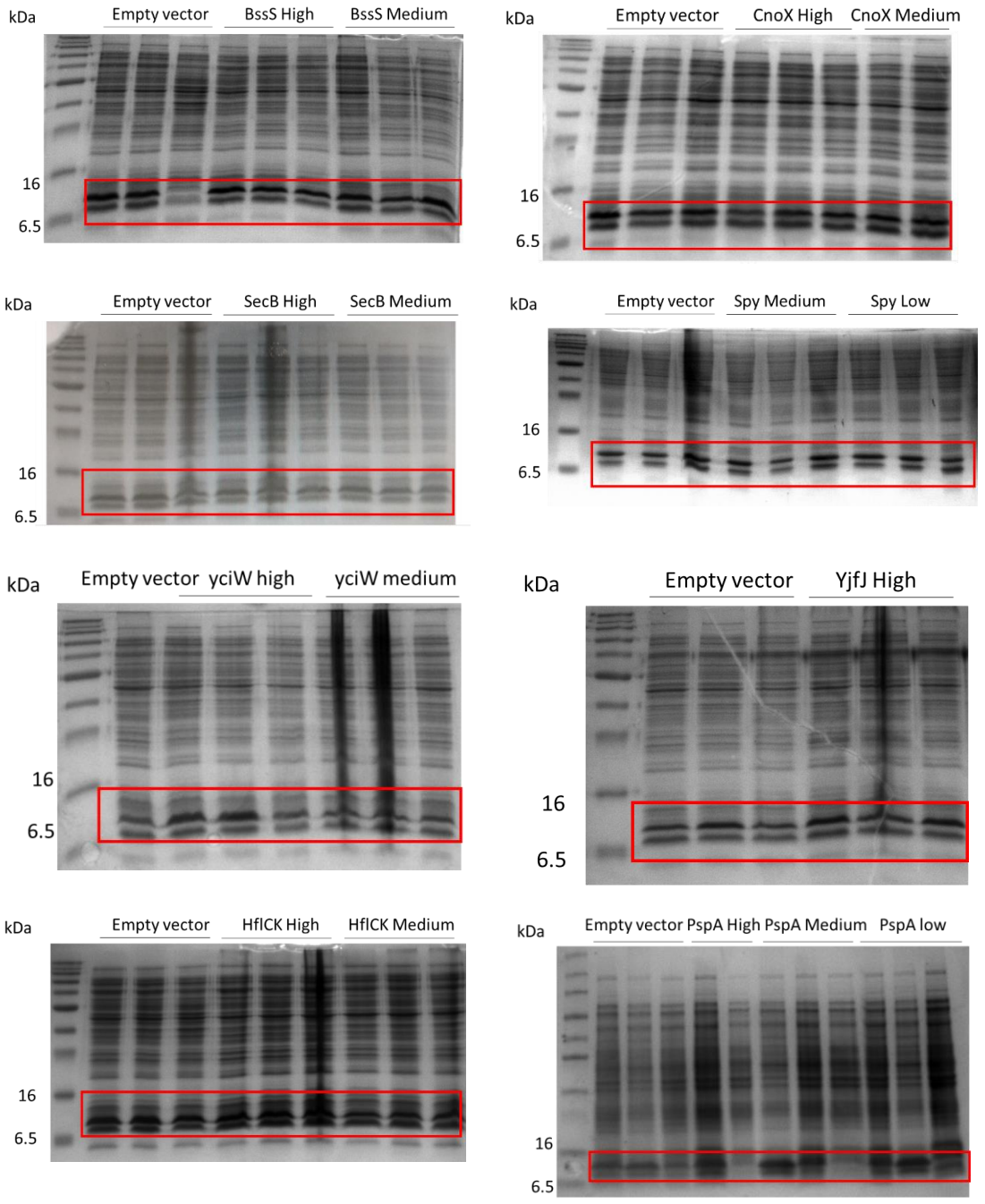
Appendix 27. SDS-PAGE, anti-pspA western blots and growth curve of strains containing the pBAD33-pspA plasmid induced with different concentrations of arabinose. Expression of pspA was induced mid log phase with different concentrations of arabinose (0-0.2 w/v % arabinose) and samples were taken 3 hrs post induction an overnight. PspA has a molecular weight of 26 kDa and is indicated by the red arrow. A protein around 25 kDa is also present in high abundance in all samples (white arrow) which is potentially the chloramphenicol resistance marker on the pBAD33 promoter as this has a predicted molecular weight of 26 kDa. Growth curve was conducted with technical triplicates and error bars represent one standard deviation



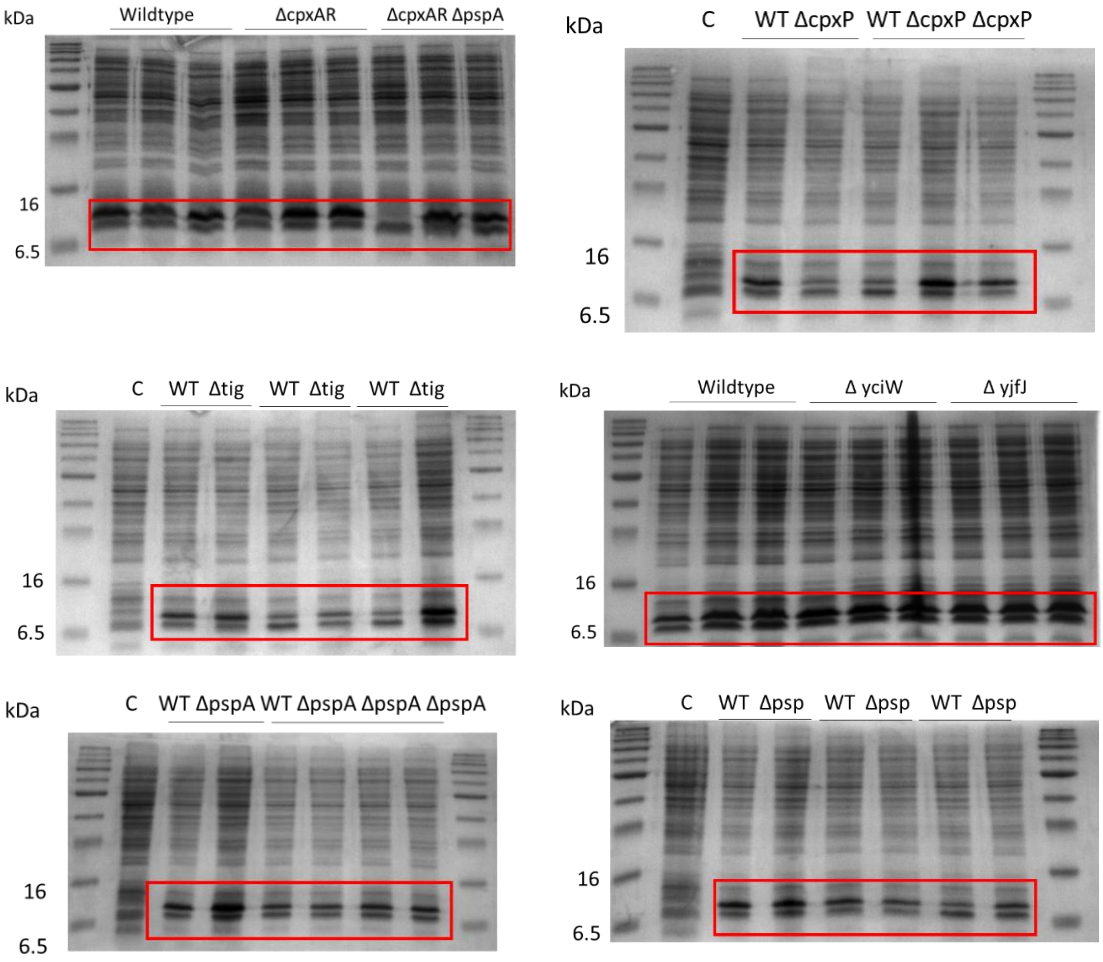
Appendix 28. SDS-PAGE analysis and anti-FLAG western blot of strains overexpressing a FLAG tagged scFv in the periplasm compared to a pre-induced control. Samples were grown in shake flasks, were induced with 0.125 mM IPTG mid log phase and samples were taken 3 hours post induction. The c-terminal flag tagged scFv has a molecular weight of 13 kDa without the signal peptide and 15 kDa with the signal peptide.



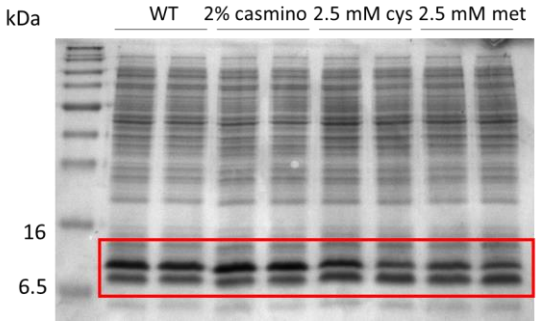
Appendix 29. SDS-PAGE of whole cell samples overexpressing scFv comparing strains overexpressing genes identified in omics data compared to an empty vector control (pAS15a_cm). Samples were grown in shake flasks, induced with 0.125 mM IPTG and samples were taken 3 hours post induction. The scFv has a molecular weight of 14 kDa with the signal peptide and 12 kDa without (secreted) and is surrounded in red. Repeats are technical replicates.



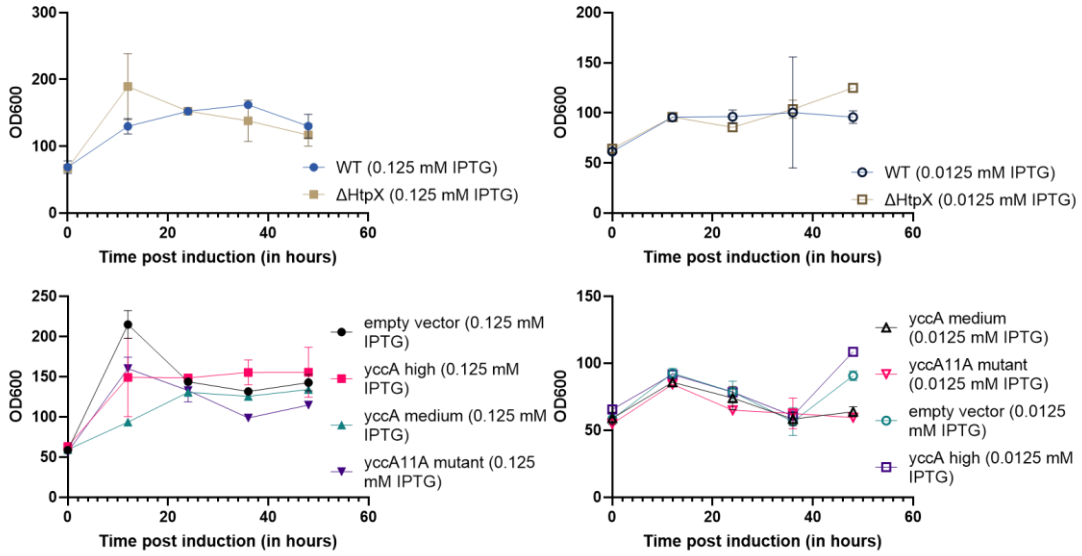
Appendix 30. SDS-PAGE of whole cell samples overexpressing scFv comparing knockout strains compared to a wildtype (WT) control. Psp- psp operon. Samples were grown in shake flasks, induced with 0.125 mM IPTG and samples were taken 3 hours post induction. The scFv has a molecular weight of 14 kDa with the signal peptide and 12 kDa without (secreted) and is surrounded in red. A pre-induced sample was used as a negative control (C). Biological replicates were taken for each sample and are grouped.



Appendix 31. SDS-PAGE of whole cell samples overexpressing scFv normal media (WT) and media supplemented with 2% (w/v) casamino acids, 2.5 mM cysteine and 2.5 mM methionine. Samples were grown in shake flasks, induced with 0.125 mM IPTG and samples were taken 3 hours post induction. The scFv has a molecular weight of 14 kDa with the signal peptide and 12 kDa without (secreted) and is surrounded in red. Replicates represent technical repeats.

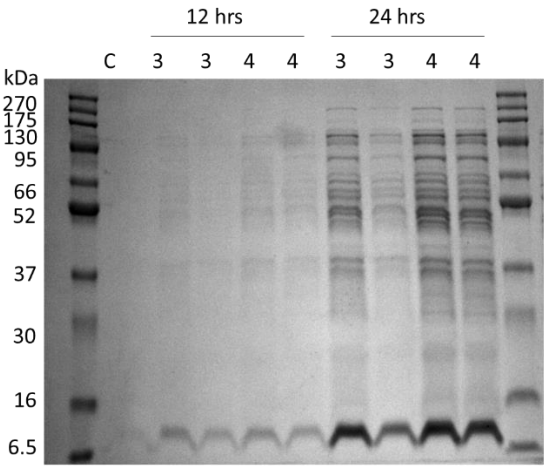


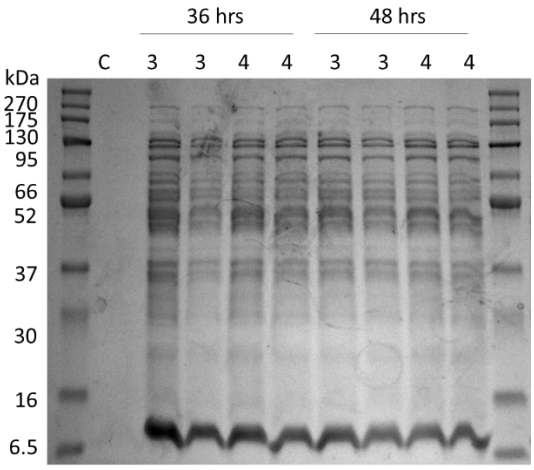
Appendix 32. Growth curves monitoring growth of engineered strains when expressing the scFv in the periplasm measured using OD_{600}. Samples were grown under industrially relevant Ambr fermentation conditions and induced with either 0.125 or 0.0125 mM IPTG.



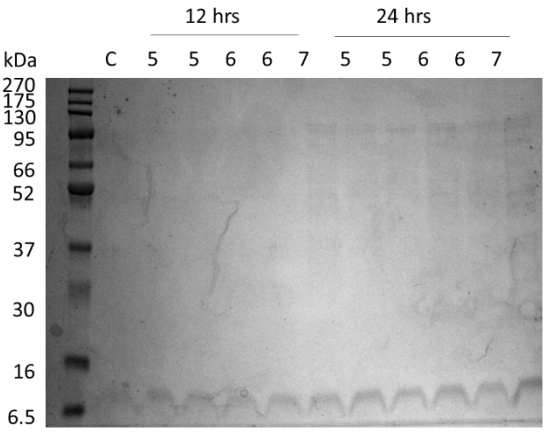
Appendix 33. SDS-PAGE of supernatants from high cell density cultures of engineered mutants expressing scFv in the periplasm (scFv+sp) by inducing expression with 0.125 mM IPTG or 0.0125 mM IPTG, 12, 24, 36 and 48 hours post induction. Most bioreactors were run in duplicate unless otherwise specified. Wildtype and empty vector samples were used as a control. Concentration of supernatants was adjusted based on the bacterial concentration (OD_{600}). A pre-induced sample was used as a negative control (C). One sample was added to the wrong lane and therefore has been crossed out. Samples were analysed on 12% gels.

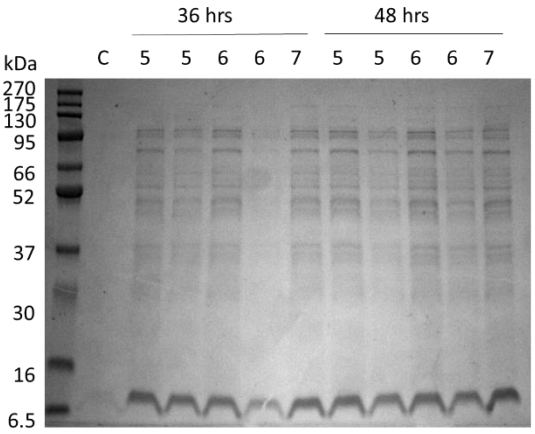
.



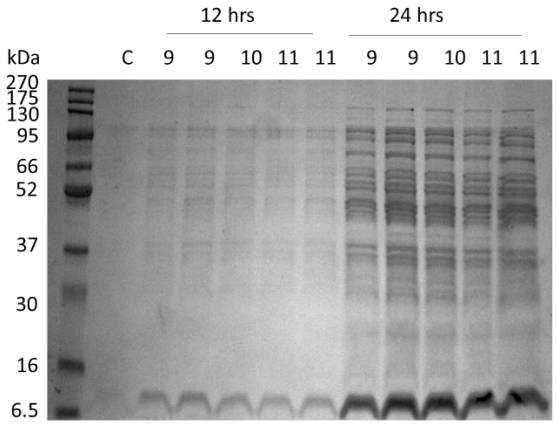


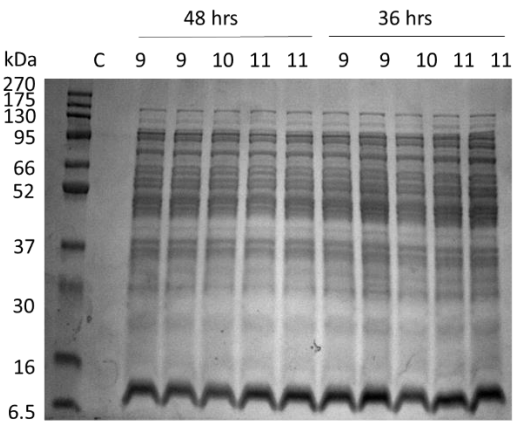
C	Pre induced control
3	WT scFv+sp 0.0125 mM IPTG (low)
4	ΔHtpX scFv+sp 0.0125 mM IPTG (low)



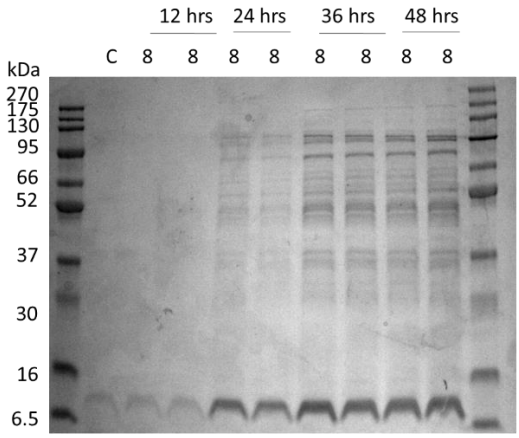


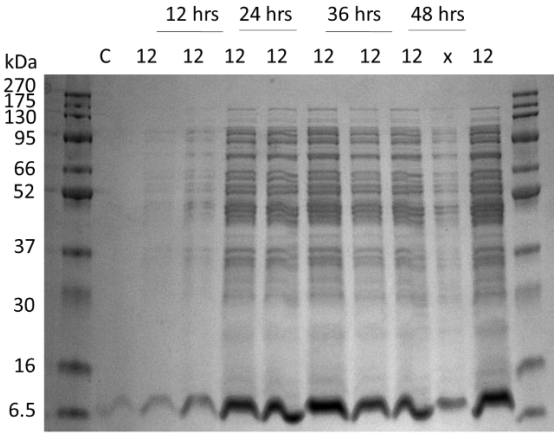
C	Pre induced control
5	Empty vector scFv+sp 0.125 mM IPTG (high)
6	yccA high scFv+sp 0.125 mM IPTG (high)
7	yccA medium scFv+sp 0.125 mM IPTG (high)





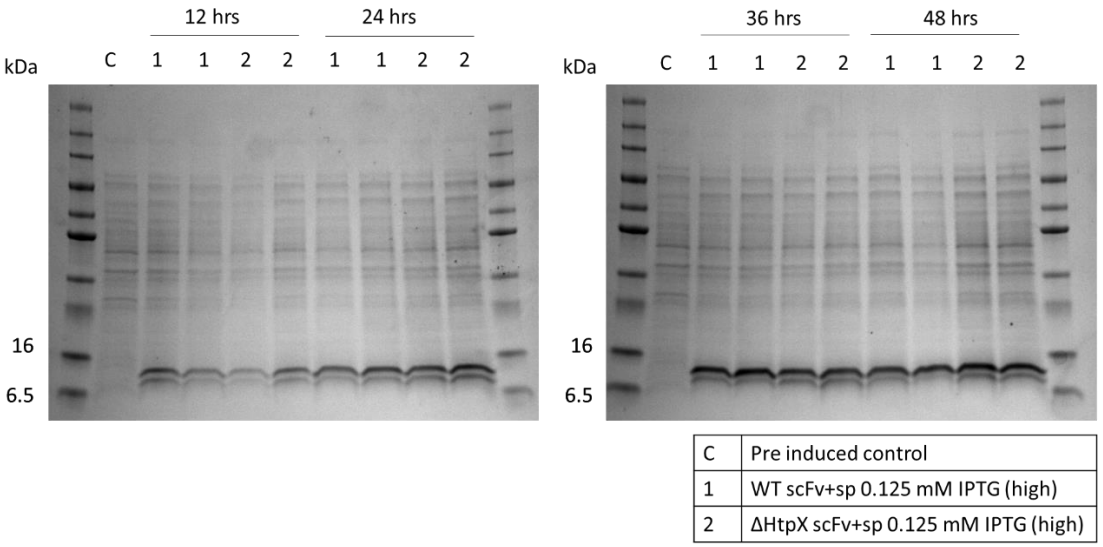
C	Pre induced control
9	Empty vector scFv+sp 0.0125 mM IPTG (low)
10	yccA high scFv+sp 0.0125 mM IPTG (low)
11	yccA medium scFv+sp 0.0125 mM IPTG (low)



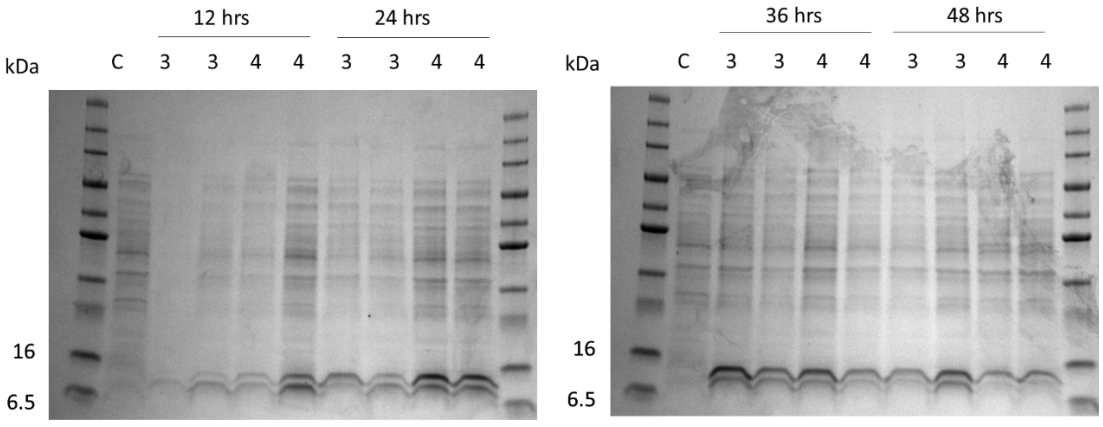


C	Pre induced control
8	yccA11a (mutant) scFv+sp 0.125 mM IPTG (high)
12	yccA11a (mutant) scFv+sp 0.0125 mM IPTG (low)

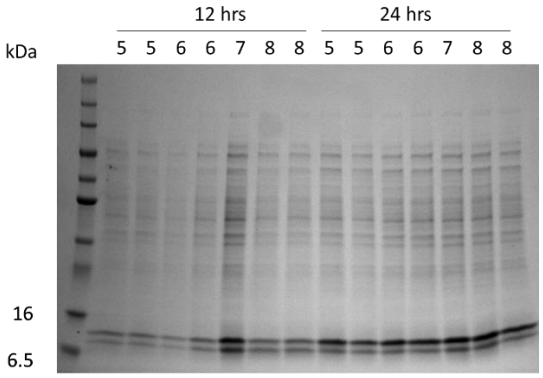
Appendix 34. SDS-PAGE of whole cells from high cell density cultures of engineered mutants expressing scFv in the periplasm (scFv+sp) by inducing expression with 0.125 mM IPTG or 0.0125 mM IPTG, 12, 24, 36 and 48 hours post induction. Most bioreactors were run in duplicate unless otherwise specified. Wildtype and empty vector samples were used as a control. Pellets were adjusted based on bacterial concentration (OD_{600}). A pre-induced sample was used as a negative control (C). Samples were analysed on 4-20% gels.

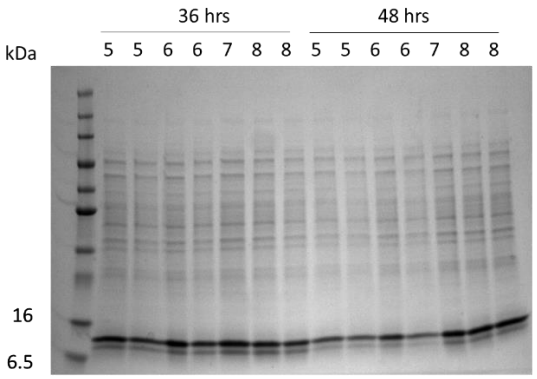


C	Pre induced control
1	WT scFv+sp 0.125 mM IPTG (high)
2	ΔHtpX scFv+sp 0.125 mM IPTG (high)

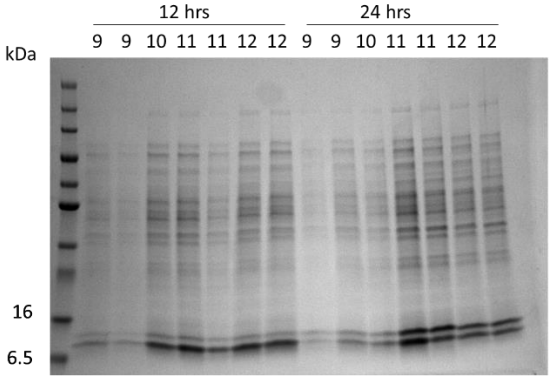


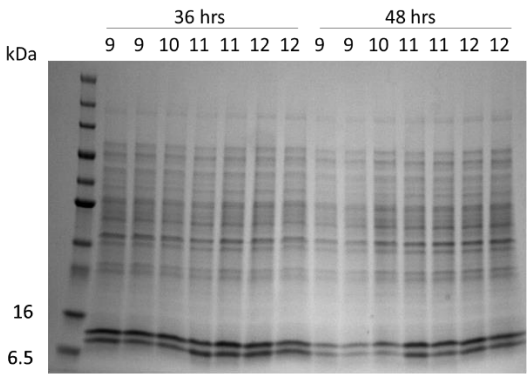
C	Pre induced control
3	WT scFv+sp 0.0125 mM IPTG (low)
4	ΔHtpX scFv+sp 0.0125 mM IPTG (low)





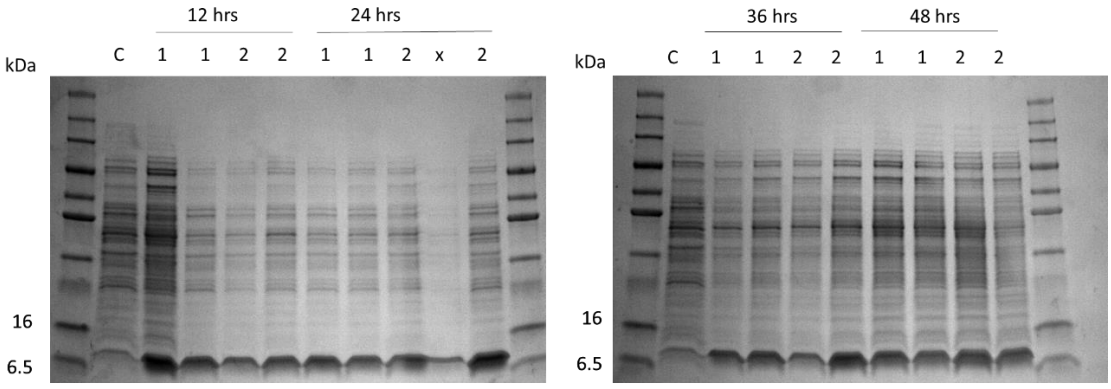
5	Empty vector scFv+sp 0.125 mM IPTG (high)
6	yccA high scFv+sp 0.125 mM IPTG (high)
7	yccA medium scFv+sp 0.125 mM IPTG (high)
8	yccA11A (mutant) medium scFv+sp 0.125 mM IPTG (high)



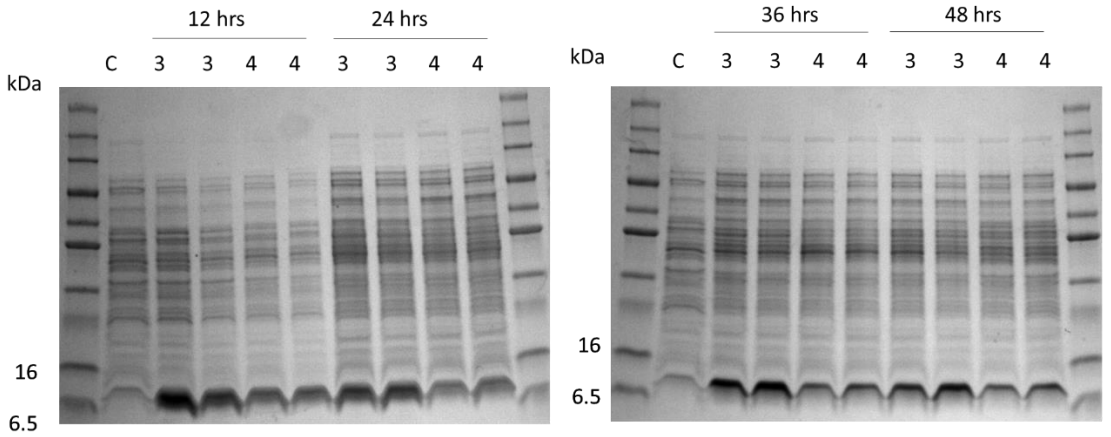


9	Empty vector scFv+sp 0.0125 mM IPTG (low)
10	yccA high scFv+sp 0.0125 mM IPTG (low)
11	yccA medium scFv+sp 0.0125 mM IPTG (low)
12	yccA11A (mutant) medium scFv+sp 0.0125 mM IPTG (low)

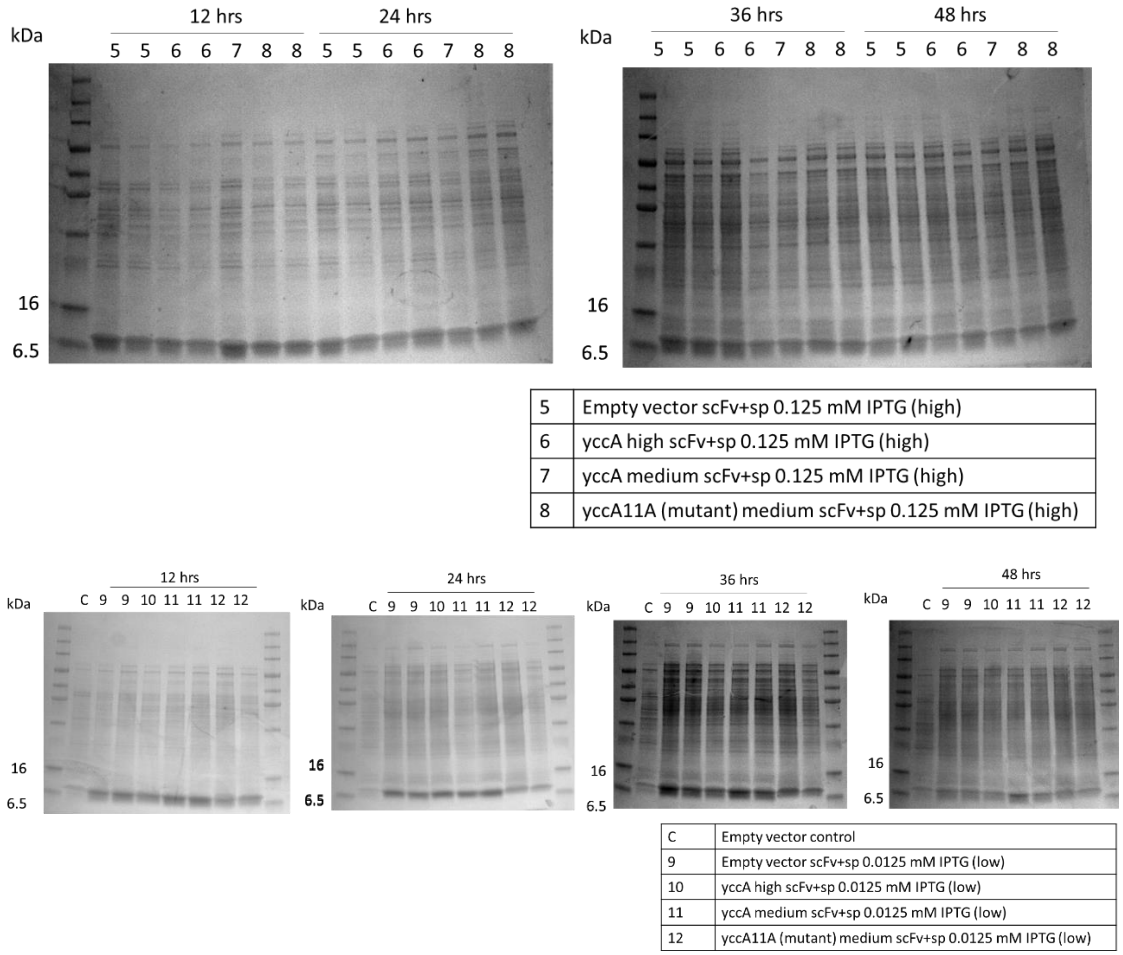
Appendix 35. SDS-PAGE of periplasmic fraction from high cell density cultures of engineered mutants expressing scFv in the periplasm (scFv+sp) by inducing expression with 0.125 mM IPTG or 0.0125 mM IPTG, 12, 24, 36 and 48 hours post induction. Most bioreactors were run in duplicate unless otherwise specified. Wildtype and empty vector samples were used as a control. Periplasmic fractions were adjusted to bacterial concentration (OD_{600}). A pre-induced sample was used as a negative control (C). Samples were analysed on 4-20% gels.



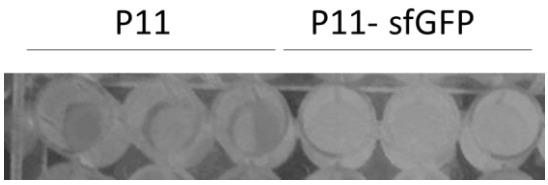
C	Pre induced control
1	WT scFv+sp 0.125 mM IPTG (high)
2	ΔHtpX scFv+sp 0.125 mM IPTG (high)



C	Pre induced control
3	WT scFv+sp 0.0125 mM IPTG (low)
4	ΔHtpX scFv+sp 0.0125 mM IPTG (low)



Appendix 36. Fluorescence of samples overexpressing sfGFP, compared to an empty vector control P11. 200 µL of bacterial cultures grown in shake flasks and induced with 0.125 mM IPTG mid log phase were places on a UV box to measure fluorescence. Biological triplicates were taken for each sample.



Appendix 37. SDS-PAGE and ponceau stain to test for equal loading of lanes in western blots.